JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44: 1949

NUMBERS 245–248

Published Quarterly by the

AMERICAN STATISTICAL ASSOCIATION

WASHINGTON, D. C.

1949

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44 MARCH 1949 NUMBER 245

ARTICLES

BOOK REVIEWS

A cumulative Index to Volumes 1–34, 1888–1939, and Annual Indexes thereafter, may be obtained from the Office of the Secretary of the AMERICAN STATISTICAL ASSOCIATION.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 245 MARCH 1949 *Volume 44*

ON A UNIQUE FEATURE OF STATISTICS*

GEORGE W. SNEDECOR
Professor of Statistics, Iowa State College

IN UNDERTAKING the presidency of the American Statistical Association, my chief purpose was to do my part toward raising the standards of our profession. I soon found it necessary to clarify my ideas about the nature of those standards I hoped to raise. Statistics is a sprawling subject covering loosely the collection of observational data, the summarization of these data, the drawing of conclusions based upon them, and pertinent mathematical theory. These processes and theories are the common property of many disciplines. Is there any unique feature that distinguishes the professional statistician from his fellows? If so, it should be the foundation on which standards are set up.

It is immediately clear that if there is anything that characterizes the professional statistician this thing changes in time. The earliest preoccupation of statisticians was with military and economic affairs of the state—human and material resources for making war, and the spoils of a successful campaign. Much later came a long period in which statisticians, in the words of R. A. Fisher, "appear to have had no other aim than to ascertain aggregate or average values." During this period, the theory of probability was extensively developed but its impact on statistical thinking was somewhat superficial. The present era of statistics is characterized by the emphasis on variation, notably sampling variation. Variation is interesting not only in itself as a well-nigh universal phenomenon, but more especially as one source of the uncertainty in all inductive reasoning. It seems that in our own time, the professional statistician's peculiar function is to develop and publicize the implications of variation. The future I do not pretend to know; but the consequences of variation, in the broad sense of uncertain in-

* Presidential address delivered at the 108th Annual Meeting of the American Statistical Association on December 28, 1948.

ference do not seem to have been entirely worked out. It is reasonable to believe, then, that both at present and in the foreseeable future, the professional statistician's most useful contribution to science is in the theory and practice of uncertain inference. Dr. Flood has put it rather strikingly in this fashion: "The professional statistician reduces data to numerical form and uses them to measure the fallibility of a conclusion where fallibility is estimated exclusively from the data in hand."

The last clause of this definition should be emphasized in advance of further discussion. All scientists make judgments about fallibility. They scrutinize their data with care; conclusions are checked against the theories prevailing in their fields; if there is a reasonable doubt about the conclusion, additional data are collected. It is only after the investigator has satisfied himself, with some high degree of assurance, that his conclusions are valid that he releases them for the critical observation of his colleagues. The distinction between the professional statistician and his fellow scientists is that the statistician evaluates the uncertainty of the conclusion by use of the data themselves, the evaluation being in the form of an exact statement of probability.

Two other facts should be observed. The first is that scientists are plagued with many variables beside those that can be reduced to measurement by the laws of chance. Inaccuracies of various kinds may creep in; it is only lack of precision, ordinarily numeralized in experimental or sampling error, that can (through appropriate conduct of the investigation) be measured. The inaccuracies may invalidate the conclusion despite the fact that the statistical measure of fallibility indicates a high degree of confidence in it. From this, one may decide that the contribution of statistics, in the restrictive sense in which I am using the term, is of minor or even negligible utility. In what follows, I hope to show that such is not the case, at least so far as my experience and observation are evidential.

The second fact to be observed is that the professional statistician and the investigator in economics, biology, engineering, etc. are usually the same person. It is merely for convenience that I mention them separately. Researchers in many fields have seized upon the statistical devices for measuring uncertainty, so that I include them in my definition of professional statistician. My thesis is that the characteristic which distinguishes the present-day professional statistician is his interest and skill in the measurement of the fallibility of conclusions.

The layman, I am sure, would be surprised by such a statement. He is accustomed to the trappings of statistics rather than to the essence

of it. To him statistics is symbolized by long rows of tedious figures and their display in tables and charts. Even job specifications for statisticians are commonly limited to the arithmetical processes of calculating averages, correlation coefficients, trends and probable errors. These are all useful procedures but they can be carried on as well by a clerk as by the professional—often, indeed, better. The professional statistician, whatever his other necessary qualifications, would seem to be set off from the layman by his habitual awareness of the fallibility of conclusions based upon data. It is not my purpose to advance further arguments for the thesis that statistics has this unique function, but merely to assume that it has and to discuss some of the consequences.

First, let us look at the field of experimental science. In the familiar sequence of the scientific method—hypothesis, experiment, conclusion—the part which is peculiarly statistical is the conclusion. This involves a judgment about the fallibility of that " . . . logically hazardous process—the process of generalizing from particular results." I am quoting from Mood's introduction of his "Theory of Statistics." "The broad problem of statistical inference is to provide measures of the uncertainty of conclusions drawn from experimental data." One highly developed measure of uncertainty is the statistical test of hypothesis; this exemplifies the statistical part of the scientific method.

Were the professional statistician to take no further interest in the procedures of the scientific method he would be fulfilling his essential share in them by evaluating fallibility, but he would fall far short of realizing his full usefulness to his fellow scientists. It is not until late in the sequence that the statistical part of the method finds its place. At this stage it is often discovered that faults in the design of the experiment make the measurement of fallibility either unnecessarily difficult or wholly impossible; it is not unusual to find too late that the quantity of data furnished by the experiment is inadequate to detect the effects in question; and it may be evident that the experiment by change of design could have been made more sensitive, with consequent saving of effort and money. This means that the professional statistician may not only facilitate his own job at the end but may increase the efficiency of the experiment by modifying its design and may, indeed, rescue it from failure by estimating its required size; all this by anticipating his own peculiar part in the conclusion. To me it is astonishing as well as gratifying that the professional statistician, in order to perform effectively his small but indispensable part in the scientific method, has been impelled to inspect the whole structure and has brought about substantial strengthening in many of its members.

Turning next to the field of science in which the survey instead of the experiment is the device used, we find statistics occupying somewhat the same unique position. The objective of the survey may be either to get information about some hypothesis or to estimate one or more parameters of the population. A survey is planned and executed in order to get the necessary evidence. The conclusion, in so far as it is based on the data, is inductive in nature and is subject to uncertainty. It may be looked upon as the professional statistician's particular business to evaluate this uncertainty.

As in experimental science, the professional statistician can enhance his usefulness by helping with the design of the survey. He can recommend designs that will furnish appropriate estimates of both position and scale; he can help choose the design that will be as efficient as is profitable; and he can specify the size of the sample that, with a designated probability, may be expected to yield a satisfactorily small measure of the fallibility of the conclusion regardless of what this conclusion may turn out to be.

But in this branch of science, the professional statistician is called upon to make more extensive contributions than those required in experimentation. Not only must he concern himself with precision but more especially with conditions affecting accuracy. So far as I can judge, the majority of the surveys now in operation have sources of inaccuracy not amenable to measurement by means of the data obtained. Restriction of the sampling to regions non-randomly chosen and the purposive selection of respondents are inherent causes of inaccuracy in the usual type of quota sampling. These causes easily could be remedied. But the professional statistician cannot stop when the relatively simple procedures of sampling are improved. He will then have to join other scientists in their attack on the really tough problems of schedule construction, selection and training of interviewers, and the little understood relations between interviewer and respondent. Although these problems are psychologic, social and economic, they affect the probabilities which control the measurement of uncertainty and therefore fall within the purview of the professional statistician's interest.

Why do professional statisticians in surveys have graver responsibilities than those in the experimental sciences? One reason is that the experiment is an older and better developed instrument than the survey, requiring less extensive and less obvious improvements. Another is that in the main experimenters are better trained for their work than are surveyors, and are heirs to a tradition of severe self-discipline.

Operators of surveys are only beginning to feel the need for examining their procedures: the embarrassment of the pollers in the recent election will, I hope, emphasize the necessity for higher standards among all samplers. If not, increasing loss of confidence by the public is sure to ensue. A third reason for the heavier responsibilities of professional statisticians in surveys is that controls are more difficult for investigators who work with human material—*homo sapiens* is notoriously a difficult experimental animal. It may be years or even decades before the professional statistician's part in the survey becomes so specialized as it now is in the experiment.

Reverting next to my opening theme, it seems obvious to me that in assessing professional standing in statistics, expertness in evaluating the fallibility of conclusions should play a major role. In saying this, I am not ignoring the fact that most users of statistics will have little interest in qualifying as specialists in so narrow a branch of the subject. Statisticians (who may or may not rate as professionals) have astonishingly varied activities. The collection of data, the planning that precedes this collection, the summarizing processes that follow, the interpretation and reporting of the results— these are preoccupations of thousands of us. Other thousands, doubtless the majority, have only an incidental interest in professional statistics, their primary objectives being in the field of application—economics, industry, medicine, and dozens more: these usually are included in the fold of statistics because, in their own subject matter fields, they base their investigations on observational data. But all who use statistics have this in common: they are working toward conclusions based on more or less incomplete enumeration, conclusions that have the uncertainty of all induction. So, they are all concerned with the evaluation of uncertainty whether or not they specialize in this unique feature. I believe that every statistician will be more valuable in his own area if he clearly apprehends this universal characteristic of his material and that his professional competence will increase with his expertness in evaluating the fallibility of conclusions based upon such material.

It is plain that I make a clear distinction between professional statisticians and statisticians in fields using statistics as a tool, statisticians who may have no proficiency in measuring uncertainty. These latter must of necessity make judgments about risk, but they often do this successfully without actual evaluation. They may be top-flight scientists or administrators but may never subject their probabilities explicitly to measurement. It is the measurement of the uncertainties of conclusions that distinguishes the professional statistician and which

makes him useful to other professionals in the various fields of application.

Professional statisticians may or may not be mathematicians. The more mathematics the better, but it is not essential. Of course, the mathematical statistician must develop the techniques of measurement and must carefully describe the conditions of their applicability. If unusual conditions are met, he must be called upon to devise appropriate new techniques. The non-mathematical professional statistician must gain experience in the subject-matter fields. He cannot assuredly evaluate the uncertainty of conclusions unless he is intimately acquainted with the uncertainties in the data which he uses for his measurements.

To some, on first thought, it may appear that I am suggesting unnecessarily rigorous standards. After more careful consideration they will agree, I think that this is not so. There is nothing essentially difficult in the idea of variation and its consequences; I have found that students in a first course in statistics easily grasp the concepts. The idea is certainly not new though it has received increasing emphasis during the last thirty or forty years. Actually I am making the modest suggestion that professional attainment in statistics be gauged by attitudes toward statistical thinking of the present rather than of the past. It seems to me that up-to-dateness is a minimum standard for professionalism in any field. In my thinking, standards in professional statistics must be based, at least in part, on modern developments in the subject; they must include not only proficiency but preoccupation in the measurement of uncertainty.

Let us now consider the application of my thesis to education. Fundamentally I am a teacher. I think my chief contribution to statistics is the training of hundreds of budding scientists in the straight and narrow way of uncertain inference. Until recently my field has been a narrow one, limited almost entirely to the statistics of biological experimentation. Only during the last year have I begun the development of a course in elementary statistics with broad cultural objectives. This field seemingly is without limits. The unique feature of statistics, the evaluation of risk, is part of the daily and hourly living of every one of us. Uncertainty envelops us, and success or failure in life is the summation of myriads of decisions as to which is the least hazardous course. It would seem that one or more courses in statistics would be part of every student's training: but, as Dr. Walker said in her presidential address of 1944, "I have never heard of a liberal arts college that undertook to explain to its students the stochastic nature of the universe in

which they live and move and have their being." Why is this? I fear it is because we, as teachers, have failed to make the subject vital and convincing. I am afraid we have emphasized the calculational and graphical devices rather than the essential nature of the subject. Instead of devoting our energies to bringing the student into harmony with his physical, biological and economic environment of variation and uncertainty, we have bored him with another course in arithmetic and algebra from which meaning is largely omitted. The nature of decisions based on probability, experience in sampling together with the concomitant risk in drawing conclusions, the fundamentals of our great cooperation in insurance, the social implications of betting—these are a few of the numerous facts of life that should form the structure of our courses in statistics. I believe that if, during the past fifty years, a realistic, living statistics had been taught, the subject would now be considered indispensable by most of our college administrators.

I think it an auspicious sign that a section on business statistics is being considered during this annual meeting. Business and statistics are blood brothers in that risk is basic in both. Yet most of our instruction in business statistics either ignores this common heritage or touches upon it vaguely in a chapter on sampling tucked away in the latter part of the book. About the only risk the student seems to be made conscious of is the risk of a mistake in arithmetic. Is it too much to hope that some of our forward-looking business executives take the initiative in advocating the elimination of unrealistic courses in statistics from our curricula and the substitution of functional courses in their stead? I judge this could be done in half a dozen years by an energetic organization with adequate resources. After all, business is a consumer of the college product and is in a commanding position to insist on quality control of the output.

The lack of professional standards for teachers of statistics in our colleges and universities is an astonishing feature of our times. In fields other than statistics, even an instructor is often required to have the doctor's degree in the subject matter of the department; the bachelor's degree in an almost universal minimum requirement. Yet how many teachers of statistics have been graduated from a curriculum in statistics? It apparently never occurs to the head of a department of education, for example, to ask a prospective teacher of statistics about his degrees in statistics. Most of us are graduates of such departments as economics, mathematics, business or psychology. Our academic training in statistics may have been no more extensive than the courses we now teach. Personally, we are not to blame for this because in our

generations there were no curricula in statistics. Even at present they are distressingly few. But we as teachers should be aggressively dissatisfied with such a condition. We should work for the establishment of departments of statistics and should each strive to improve his own professional standing. We should resolve that the next generation of teachers shall have advantages not available to us. Our Section on the Training of Statisticians has the glorious opportunity of leading in this high endeavor.

What are the implications of my thesis for the American Statistical Association? Under our new constitution we have abandoned our preoccupation with any one subject-matter field and have volunteered our services as a focus of statistics among them all. In this capacity, we were asked to participate in the work of the Hoover Commission, suggesting desirable reorganization in the statistical agencies of the government. We are joining the Social Science Research Council in reviewing the recent election predictions of three of the most prominent polling organizations. Two government bureaus have asked criticism and suggestions for their programs. In meeting such responsibilities, our commission on Statistical Standards and Organizations seems destined to wield a powerful influence in statistical affairs. These things can be done only because we have in our membership professional statisticians of the highest competence. To maintain and enlarge this leadership, we must attract to our ranks other able statisticians from all subject-matter fields, professional statisticians whose skills include that unique function of modern statistics, the measurement of the uncertainty of inductive conclusions.

AN ATTEMPT TO GET THE "NOT AT HOMES" INTO THE SAMPLE WITHOUT CALLBACKS

Alfred Politz
and
Willard Simmons

PART I

This paper describes a plan for eliminating the need for callbacks. Each person in the sample is visited only once. From each person interviewed information is obtained as to whether or not he was at home on specific instances, including the instance of the interview, which permits an estimate of the proportion of time he is at home during the interviewing hours. Questionnaires are divided into e.g. 6 groups according to the estimated proportion of time persons in each group are at home, viz., 1/6, 2/6, · · · , 6/6 of the time. The sample estimate, for any variable under study, is produced by weighting the results for each group by the reciprocal of the estimated per cent that persons are at home. It is shown that under certain conditions this estimate is unbiased and the variance of the estimate is obtained. A numerical comparison is made between this plan and the usual method of calling back.

Many individuals are not available for an investigation on the first visit because they are not at home when the interviewer calls on them. These cases are often referred to as the "not at homes." Depending on the time when interviews are feasible and on the kind of individuals under question, the percentage of "not at homes" usually varies between 30 and 60. The "not at homes" thereby constitute a factor of extreme importance. The simplest theoretical device for the completion of the sample consists of revisiting, again and again, the homes where a certain individual was not found on the first call, until the particular individual is found. These callbacks are spread thinner and burdened with longer travel time than first-call interviews. The second visit is more expensive than the first visit, the third visit is more expensive than the second one. The economic burden increases with subsequent calls to the point that a certain percentage of attempted interviews usually is considered unobtainable. The increased costs per

information unit derived from callbacks make it, in most cases, advisable not to attempt revisiting all the "not at homes" of the primary sample, but to revisit only a sub-sample of it.[1] While sub-sampling increases the sampling error, it still need not introduce biases. The biases start only where revisiting of "still not at homes" stops. With the callbacks as a major source of expense in unbiased population samples, it has been worthwhile to study the possibility of circumventing the need for them altogether. During the past three years, we have developed a plan for eliminating the need for callbacks and several experiments have been made applying this plan to market surveys.[2]

The first step may be a review of the meaning of the "not at homes." 1) If the survey is concerned with items open to observation within the household or with items about which nearly every adult member of the household can give information, it usually suffices to design a sample of households. An investigation of a household in the sample becomes impossible if nobody is at home (more accurately, no adult is at home) at the time when the interviewer rings the doorbell. 2) If the investigation is concerned with problems where an individual reports about himself (buying habits, opinions on social and political issues, taste preferences, etc.), a sample of individuals is designed. Under these circumstances, it no longer suffices that somebody (some adult) is at home. A *particular* individual has to be found. If this particular individual is not at home when the interviewer rings the doorbell, the information cannot be obtained. This paper is concerned solely with samples of individuals. It is obvious that the not-at-home rate in samples of individuals is higher than the not-at-home rate in samples of households.

If several callbacks are made in order to reach individuals not found at home on the first visit, the assumption is maintained that the individual *not at home at one instance*, will be *at home some time* during those hours when personal visits are possible. Individuals who are at home only during night hours, let's say from 10 P.M. to 8 A.M., drop out of almost every sample. By leaving such inaccessible extremes aside, we may say that people who are not at home at the first visit but can be found in a second, third, fourth, fifth or sixth visit, are persons who

---

[1] A plan for sub-sampling non-responses to mail questionnaires is discussed by Morris H. Hansen and William N. Hurwitz, *The Problem of Non-Response in Sample Surveys*, Journal of the American Statistical Association, December, 1946, page 517. The principles set forth for determining the size of the original sample and the size of the sub-sample of non-responses to maximize sampling efficiency for a given cost are applicable to the "not at home" problem.

[2] It has recently been brought to the authors' attention that a somewhat similar plan was proposed independently by H. O. Hartley before the Royal Statistical Society. This proposal was made in commenting upon a paper by F. Yates, *A Review of Recent Developments in Sampling and Sampling Surveys*, Journal of the Royal Statistical Society, Vol. CIX, Part I, 1946, page 37.

stay away from home more often by varying degrees. The average frequency of staying away from home among the second call respondents is higher than among the first call respondents; the average frequency of staying away from home among the third call respondents is higher than among the second call respondents, and so forth. Because of the fact that some people are away from home more often, it becomes necessary on the average to visit their homes more often before they can be found at home. But at one time if callbacks are continued indefinitely, they actually are found by the interviewer. Let's say an interviewer finds the respondent, Mr. Smith, at the occasion of the third call at 8 o'clock in the evening on Wednesday. If the survey schedule and interviewing assignment had acidentally brought the interviewer to the home of Mr. Smith in the first place at 8 o'clock on Wednesday night, he would have been found at home at the first visit. Mr. Smith then never would have belonged to the group of "not at homes."

This may make it obvious that every sample of "at homes" must include "potential not at homes." To put it more accurately, every set of first call interviews of timing A must include respondents who are "not at homes" in another set of interviews of timing B. It must include respondents who are not at home in timing C and it must include respondents who are not at home in timing B and timing C. Statistically, therefore, it must be possible to reconstruct from a present "at home" sample, past samples of "at homes" and "not at homes," if: (a) respondents provide information on their past "at home" performance, and (b) if the individuals in the present "at home" sample are visited at times chosen at random.

Consider, for example, the following three groups, among which all individuals in the population are distributed: 1) those who are at home, on the average, 20% of the time, 2) 50% of the time, and 3) 80% of the time. If the time of visits is determined at random, we would expect to find on the first call about 20% of group (1), 50% of group (2), and 80 of group (3). Now if each person in the sample can only be identified with the group to which he belongs, a correction for the under-representation of each group is clearly indicated. Since only about one-fifth of the persons in the first group are interviewed, this group is assigned a weight of 5. Likewise, the second group receives a weight of 2 because only about half of the persons in this group are found at home, while the third group receives the weight of 1.25. This weighting, of course, does not completely eliminate the bias, for it takes into account only three arbitrarily defined groups. On the other hand, the bias must be reduced because the weighting has at least partially compensated for

the under-representation of persons frequently away from home.

The number of such groups, however, need not be restricted to three. With obvious modifications, the above example is applicable to any number of such groups into which the population might be divided, where each group contains persons who are at home any part of the time during which interviewing is in progress. Consider the limiting case in which there are as many such groups as there are persons in the population; that is, each person is a group of one. If it were possible to assign to each individual found at home, a weight equal to the reciprocal of the per cent of time he spends at home, the not-at-home bias would vanish.[3] Such weighting would completely compensate for the under-representation of persons frequently away from home.

Estimating the Per Cent of Time Respondents Are At Home

While it is hardly possible to find out the exact percentage of the time an individual is at home, it has proven feasible to estimate this percentage from information obtained by direct questions to the respondent himself. The problem of phrasing such questions has naturally received considerable attention and experimentation. Any such question as, "Are you usually at home in the evening?", of course, is valueless. A more specific question, such as, "How many nights out of the last five, were you at home?", is much better, but is still subject to two objections: 1) the respondent is likely to answer without thinking back over his activities on the previous five evenings, and 2) no provision is made for the respondent who was at home during part of an evening and away for the remainder of the evening. While further improvements are possible, the following questions have proven to be satisfactory: 1) Would you mind telling me whether or not you happened to be at home last night at just this time? 2) How about the night before last at this time? 3) How about Wednesday night? 4) How about Tuesday night? 5) Monday night? These questions relate specifically to interviews conducted on Saturday and the particular days of the week mentioned in questions 3, 4 and 5 are changed, as appropriate for interviews conducted on other nights. To alleviate any possible resentment at the personal nature of the inquiry, interviewers find it helpful to preface the questions by some statement like, "We are also interested in finding out how often people go out in the evening at various times and on various days of the week. I wonder if you would mind telling me . . . etc."

[3] See Part II, page 22, footnote 10.

It is important to decide upon an optimum number of nights about which inquiries should be made, bearing in mind the limitations of respondents' memories and willingness to cooperate, as well as the obvious advantage of having information for as many nights as possible. Experience with this technique in several field operations has led to the acceptance of information for five previous nights as perhaps the most suitable, where evening interviews are conducted Monday through Saturday. Since the intreviewer obtains information about one night by observations, this makes a total of six nights upon which an estimate may be based of the actual per cent of time respondents are at home. A clear advantage lies in the fact that six nights coincide with a complete week of interviewing, and the effects of any tendency of respondents to go out more frequently on certain nights of the week is eliminated. Experience with this plan has indicated that respondents are almost always able and willing to give answers for five nights previous. This is borne out by the extremely small number of non-responses and "don't knows" to these questions. Because the questions relate to an individual's personal activities, it is the type of information respondents do know, and are not reluctant to impart.

Sample Projections Based on Unbiased Estimates of the Time Respondents are At Home

Information concerning the number of nights each respondent is at home, out of six specified nights at a particular time, provides an unbiased estimate of the actual per cent of the time each respondent is at home.[4] This information makes it possible to divide the respondents into six groups, each having a weight depending upon the proportion of such individuals expected to be found at home, as follows:

(1) Group	(2) Estimated proportion of time spent at home	(3) Weight
1	1/6	6.0
2	2/6	3 0
3	3/6	2 0
4	4/6	1 5
5	5/6	1 2
6	6/6	1 0

The weights in column (3) are the reciprocals of the estimated proportion of the time individuals in each group are at home (column 2).

[4] See Part II, page 18.

Information obtained from interviews with individuals in each group may be multiplied by the corresponding group weight to produce an estimate for all individuals in the group, of those originally in the sample, including persons not at home. There is one group, however, in the original sample, from whom no interviews have been received; that is the group containing those persons who were not at home on any of the six nights, including and preceding the night the interviewer called at their homes. This group, of course, is comparable to the group which is not reached even after five callbacks.

If information is obtained for six nights at random, estimates based on this technique are subject to no not-at-home bias other than the bias contributed by the relatively small group who are not at home on any of six selected nights.[5] As far as bias is concerned, therefore, such an estimate is equivalent to that obtained by a sampling plan in which six calls are made when necessary to find the "not at homes." In making this comparison, however, some allowance must be made for the additional contribution to the sampling error because of failure to obtain interviews from any "not at homes." In actual practice, it has been found that this increase in sampling error seldom exceeds 2% (coefficient of variation). Moreover, for almost all situations likely to be encountered in practice, the use of this technique in combination with only one callback will produce more reliable estimates, considering both the bias and sampling error, than may be obtained from a five-callback interviewing operation which does not provide for information on previous "at home" performance.

The great expense of repeated callbacks, however, has already led to extensive sub-sampling of the "not at homes." This procedure, developed by the Census Bureau, has gained wide acceptance by scientific samplers, because it frequently produces closer estimates for the cost. In a sub-sampling operation, however, consideration must also be given to increased sampling error. The addition to sampling error, occasioned by a sub-sampling operation, is likely to be greater than the increase resulting from use of the proposed plan, even assuming an optimum allocation of interviews on first call, second call, and subsequent callbacks. In fact, the numbers of persons, discovered by inquiry, who are at home one night out of six, two nights out of six, etc., will usually correspond fairly closely with the optimum allocation of callback interviews, if this information had been available. This assumes, of course, that the optimum numbers of interviews to be obtained on successive callbacks are determined after taking into account the increased costs

[5] See Part II, page 21.

of obtaining interviews after repeated callbacks. It would seem, therefore, that for population studies, employing samples based on probability theory instead of the inevitable errors in judgment, the use of this plan will yield impressive economies of both time and money.

While this paper uses the past 5 days' performance with regard to being at home as the basis for the development of strata, workers who want to use this method may deal with situations in which a smaller number of days is justified. The decision they have to face is similar to deciding on the maximum number of callbacks to be made on "not at homes." On a special survey in which the writer had to get a measurement of the biases in a judgment sample, it was considered necessary to make up to eight callbacks. In surveys where less precision is required, three or even two callbacks are set as a maximum. It is impossible, without reference to the particular subject under study, to make a final statement about the number of callbacks necessary, or about its approximate equivalent; that is, the number of days' past performance that should be covered.

While the mathematics of the "not at home" calculation is explained in Part II, one point, of a psychological nature, may require a reference from practical experience. It is the question mentioned before as to whether respondents can report about their past at-home performances with sufficient accuracy. There is no doubt that many survey questions burden the memory and honesty of the respondent much more than the at-home question. However, it would not be good policy to take possible inaccuracies in sampling lightheartedly just because surveys as a whole have their weaknesses anyway. It is for this reason that actual field experiences should be quoted. Following small scale experiments which cleared the path, the method has been employed since the latter part of 1947 in major area surveys.

It is well known that men and women differ substantially in their at-home performance. Therefore, in a survey of the Philadelphia metropolitan area, the proportion of females and males in the population was left to discovery by the survey using the new device.

PHILADELPHIA METROPOLITAN AREA

Per cent males directly counted in at-home sample	Per cent males estimated from at-home questions	Per cent males estimated by U. S. Census, 1947
43.1%	47.8%	47.4%

The difference between the survey estimate and the Census estimate is well within the sample tolerance.

An internal check on response reliability is provided in this plan without resorting to comparisons with Census data. On the one hand, a record may be kept by interviewers of the number of persons found at home and the number not at home of all persons visited. This permits a direct estimate of the per cent of persons who are at home when the interviewer calls. On the other hand, the expected per cent of persons at home at a given time chosen at random is equivalent to the average per cent of the time all persons are at home. This may be estimated directly from the information obtained from respondents concerning the number of nights each respondent was at home out of the past six nights. This comparison between two independent estimates of the average per cent of persons at home is usually made in order to check the over-all accuracy of respondents' answers, interviewers' records and and any other source of error. The results of this check for the Chicago survey are as follows:

CHICAGO METROPOLITAN AREA

	At home per cent	Not at home per cent
Based on actual interviewers' records of number of persons visited and number found at home	61.1	38.9
Based on respondents' answers concerning the number of nights they are at home	61.5	38.5

It is advisable for anyone who may use the method in the future to maintain a direct count of the individuals not found at home, although it does not add to the information which is sought by the survey. However, the direct count provides an elegant internal statistical check on the reliability of the response to the not-at-home question, and thereby indirectly, a check on the interviewers' carefulness in dealing with the respondents.

FURTHER THEORETICAL CONSIDERATIONS REGARDING THE PLAN FOR ELIMINATING CALLBACKS

PART II

THE PLAN FOR eliminating callbacks described in Part I may be summarized briefly as follows:

1) Each person in the sample is visited once and only once at a time determined at random, considering only the periods during which interviews are to be conducted.
2) From each person interviewed, information is obtained as to whether or not he was at home at six specific instances, determined at random, including the instance of the interview, which permits an estimate of the proportion of time he is at home during interviewing hours.
3) Questionnaires are divided into six groups according to the estimated proportion of time persons in each group are at home, viz., 1/6, 2/6 · · ·, 6/6 of the time, for groups one to six, respectively.
4) The sample estimate, for any variable under study, is produced by weighting the results for each group by the reciprocal of the estimated per cent of time persons in the group are at home. Thus, the weights for groups one to six are, respectively, 6/1, 6/2 · · ·, 6/6.

Assumptions Made

The population to which the sample estimate relates is restricted to those individuals who are at home at least some time during interveiwing hours; that is, those persons who could eventually be found by callbacks during regular interviewing hours. The decision concerning the hours of interviewing is, of course, extremely important to the survey results, and the shorter the daily interviewing period, the larger the number of persons arbitrarily given no chance of being found at home.[6] For example, employed persons thus may be excluded from daytime interviews, persons attending night school may be excluded from evening interviews; whereas, neither group is excluded from an interviewing schedule including both daytime and evening hours. In the limiting

[6] Because more persons are usually at home at certain times than at other times, further consideration of the "optimum" periods during the day for interviewing may be worthwhile, possibly leading to a stratification by time of day and of the principles for optimum allocation of sample cases to such strata.

case, the excluded group consists only of those persons who are never at home, or who have no home, since interviews are theoretically possible at all hours. In the following discussion it will be convenient to consider an experiment in which each of the N persons in the population is visited one time. Because the not-at-home problem is no different for a probability sample than for an attempted total census of the entire population, this involves no loss of generality. Let us assume that interviews are obtained from all of the n persons who are found at home, i.e., no person refuses to be interviewed, and that $(N\text{-}n)$ persons are not at home when the interviewer calls. Now in effect we have a sample of n individuals in which the probability of including any person is equal to the probability that person is at home when the interviewer calls.

The random choice of a time of visiting each person is to avoid the arbitrary exclusion of persons who are never at home, for example, at the time of day and day of the week at which, otherwise, it may be decided arbitrarily to visit them.[7] When an interviewer rings a doorbell, he is sampling time. He has chosen at random one particular moment from a large number of possible moments to ascertain whether or not the respondent is at home. The chance that an individual is interviewed, therefore, is exactly equal to the per cent of the time that individual spends at home, counting only the hours during which interviews are conducted. Moreover, for each of the n individuals who do happen to be at home, the interviewer obtains a sample of five additional points in time. The questions suggested in Part I, together with the interviewer's observation at the time of his visit, provide a systematic sample from a random start of a cluster of six moments spaced twenty-four hours apart. It may be easily shown that this sample provides an unbiased estimate of the per cent of time each individual is at home. Let p_{jk} equal unity if the j^{th} person is at home at the k^{th} moment, otherwise p_{jk} equals zero. Since the moments were selected with equal probability, it follows that $\tilde{p}_j = 1/6 \sum_{k=1}^{6} p_{jk}$ is an unbiased estimate of $\bar{p}_j$, the actual per cent of time that the j^{th} individual is at home ($\bar{p}_j \doteq 1/M \sum_{k=1}^{M} p_{jk}$, where M is total number of moments of interviewing.)[8]

[7] A random choice of the time of calling on any individual may be approximated closely in actual operations without creating any severe administrative problems. It does not require, for example, that the selection of a time of visiting different individuals be independently at random. It is quite feasible to select first at random an evening on which all persons in a particular location or cluster will be visited, to order the visits for convenient travel between homes, selecting at random only the point within the cluster at which the interviewer is to begin making calls at a specified time. While not strictly meeting the requirements of randomness, this procedure results in a fairly close approximation.

[8] A moment is chosen merely for convenience. Actually any other unit of time would serve as well provided it is understood that the interviewer does not wait for respondents to return home but simply takes the time necessary to ascertain whether or not the respondent is already at home when the doorbell rings. If one prefers to consider the unit of time infinitesimal, $\tilde{p}_j$ may be shown to be unbiased by employing the integral $\int p_{jt}dt$.

We may further define $\bar{p}=1/N \sum_{j=1}^{N} \bar{p}_j$ as the average per cent of time all N persons in the population are at home.

Since information concerning the time each person in the sample is at home is obtained for six moments, it is convenient to think of the population as being divided into seven groups which correspond to the actual and potential answers regarding the number of nights they were at home. The following seven groups are, therefore, defined:

TABLE I

STRATA BASED ON THE NUMBER OF NIGHTS ON WHICH INDIVIDUALS ARE AT HOME AT A SPECIFIED MOMENT

No. of nights at at home out of six	Proportion of time at home		Size of pop.	Estimated Size of pop.	Size of sample	Pop. total	Sample total
(1)	Est.	Act.					
0	0/6	$\bar{p}_0$	N_0	$\hat{N}_0$	—	X_0	—
1	1/6	$\bar{p}_1$	N_1	$\hat{N}_1$	n_1	X_1	$\hat{X}_1$
2	2/6	$\bar{p}_2$	N_2	$\hat{N}_2$	n_2	X_2	$\hat{X}_2$
3	3/6	$\bar{p}_3$	N_3	$\hat{N}_3$	n_3	X_3	$\hat{X}_3$
4	4/6	$\bar{p}_4$	N_4	$\hat{N}_4$	n_4	X_4	$\hat{X}_4$
5	5/6	$\bar{p}_5$	N_5	$\hat{N}_5$	n_5	X_5	$\hat{X}_5$
6	6/6	$\bar{p}_6$	N_6	$\hat{N}_6$	n_6	X_6	$\hat{X}_6$
Groups one through six		$\bar{p}$	N	$\hat{N}$	n	X	$\hat{X}$
Groups zero through six		$\bar{p}_t$	N_t	—	n	X_t	—

The size of the population in each group, $N_i(i=1, 2,\cdots, 6)$ is, of course, unknown. However, the N_i are rigidly defined. For example, the N_1 persons in Group one include all persons whose answers indicated that they were at home only on the night the interviewer called and on none of the other five nights, at the specified time. In addition, the population in Group one includes all persons, not at home when the interviewer called, who could have truthfully given such an answer for the previous five nights if they had been asked the questions at the particular moment that the interviewer called at their respective homes. The zero group is made up entirely of the latter, since a person who was not home at least one night in the specified six could not fall into the sample. The seven groups are mutually exclusive and exhaustive and are sampled independently except for the zero group which is not sampled at all. The other six groups may be properly treated as strata. It is assumed that the population is large enough and that the distribution among the population of the variate $\bar{p}_j$ is such that the sample does include some persons in each of the groups one to six. It will be noted that the N_i are precisely defined only after the time of visiting

each respondent is fixed, and that a change in the schedule of calls will change the numbers and identity of the persons falling in each of the seven groups. In this sense, the N_i are variates, having a sampling distribution, which assume definite values whenever the time for visiting all respondents is selected. With these conditions in mind, consider the sample estimate ($\hat{X}$) of the population total (X) for the characteristic under study.

$$\hat{X} = 6\sum_{j=1}^{n} \frac{x_j}{i}. \tag{4}$$

Where x_j is the value of the variate under study for the j^{th} person, n is the number of persons found at home and i is the number of nights any individual is at home out of the six nights including and just preceding the night of the interview. Thus, the value of the variate for each individual in the sample is weighted by the reciprocal of the estimated per cent of time he is at home.

The usefulness of this entire technique for eliminating callbacks depends largely upon whether or not $\hat{X}$ is a "good" estimate in the sense that: (1) it is unbiased and (2) has a small sampling error for the class of population for which it is appropriate.

Mean of the Sample Estimate

It follows from equation (4) that:

$$E\hat{X} = 6E\sum_{j=1}^{n} \frac{x_j}{i} = 6\sum_{j=1}^{N_t} \bar{p}_j E_j \frac{x_j}{i}. \tag{5}$$

Where $\bar{p}_j$ is the probability that the j^{th} individual is at home when the interviewer calls, and $E_j x_j/i$ denotes the conditional expectation of x_j/i knowing that the j^{th} person is found at home and interviewed. The value of x_j/i, for each person interviewed, depends upon how many nights out of the previous five nights that person was at home. Thus we have,

$$E_j \frac{x_j}{i} = \sum_{i=1}^{6} \frac{x_j}{i} \frac{5!}{(i-1)!(6-i)!} \bar{p}_j^{\,i-1}(1-\bar{p}_j)^{6-i} \tag{6}$$

in which the coefficients of x_j/i in the six terms of the summation are, respectively, the probability that the j^{th} person was at home 0, 1, 2, $\cdots$, 5, previous nights; viz., the six terms in the expansion of $(\bar{q}_j+\bar{p}_j)^5$

where $\bar{q}_j = 1 - \bar{p}_j$. Equations (5) and (6) readily yield for $E\hat{X}$:[9]

$$E\hat{X} = \sum_{i=1}^{6} \sum_{j=1}^{N_t} P_{ij} x_j \tag{7}$$

where P_{ij} is the probability that the j^{th} individual falls in the i^{th} stratum, i.e.,

$$P_{ij} = \frac{6!}{i!(6-i)!} \bar{p}_j^{\,i}(1 - \bar{p}_j)^{6-i}$$

is the $(i+1)^{th}$ term in the expansion of $(\bar{q}_j + \bar{p}_j)^6$. Since the sum of the seven terms in this expansion equals unity, we have,

$$\sum_{i=1}^{6} P_{ij} = 1 - \bar{q}_j^{\,6}. \tag{8}$$

Making this substitution in (7), the expected value of $\hat{X}$ becomes:

$$E\hat{X} = \sum_{j=1}^{N_t} x_j(1 - \bar{q}_j^{\,6}) = \sum_{j=1}^{N_t} x_j - \sum_{j=1}^{N_t} \bar{q}_j^{\,6} x_j. \tag{9}$$

But q_j^6 is the probability that the j^{th} individual will not be at home on any of the six specified nights. The second term of (9), therefore, is the expected value of the variate for the zero group, thus,

$$E\hat{X} = \sum_{j=1}^{N_t} x_j - \sum_{j=1}^{N_t} \bar{q}_j^{\,6} x_j = EX_t - EX_0 = X. \tag{10}$$

It is clear that $\hat{X}$ is an unbiased estimate of X, the total value of the variate for all persons in the population other than those who cannot ordinarily be found at home in six visits. Although X is a constant, the X_i will vary in successive samples according to the number and identity of the individuals which fall in each stratum for a particular arrangement of the interviewers' schedule of visits. As a corollary of the above proof, however, it can be shown that $\hat{X}_i$ is an unbiased estimate of the

[9] The assumption is implied that the six moments are independently selected at random. If the particular questions suggested in Part I are used, the six moments in question are not independently selected at random, but systematically selected within a randomly selected cluster of six successive nights. An exact statement of probability would involve the intra-class correlation of the probability of an individual being at home on successive nights. However, experience has indicated that this correlation tends to be quite low or even negative because a person is more apt to stay at home on a night following a night on which he goes out. The assumption of a zero correlation is, therefore, realistic.

aggregate value of the variate for the i^{th} stratum, i.e., $E\hat{X}_i = EX_i$.[10] By definition, we have:

$$X_i = \sum_{j=1}^{N_i} x_j. \tag{11}$$

It follows directly that

$$EX_i = \sum_{j=1}^{N_i} P_{ij} x_j. \tag{12}$$

Since this expression is identical to that given within the summations ($i=1, 2 \cdots 6$) in equation (7), clearly $E\hat{X}_i = EX_i$. It is also evident that $E\hat{N}_i = EN_i$ and that $E\hat{N} = N$.

Brief attention must be given to the zero group which inevitably contributes a bias to the results of any population survey. Any estimate for this excluded group is necessarily based on an assumption, for example, that $\overline{X}_0 = \overline{X}_1$, or where the variate under study is highly correlated with the tendency to be away from home, the assumption might be that $\overline{X}_0 = a + b\bar{p}_0$, in which (a) and (b) are the constants in the regressions between (x) and (p) as determined from the sample. This latter assumption might be warranted if the zero group is large and one is estimating, for example, the number of nights each week individuals go to the movies, or listen to the radio. The regression estimate has one clear advantage: the extent of the bias because of "not-at-homes" depends upon the correlation between (p) and (x). If they

[10] A necessary condition for $\hat{X}$ and the $\hat{X}_i$ to be unbiased is the random selection of a time for each visit. If the number and identity of the N_i individuals in each stratum were determined, for example, by an arbitrary arrangement of the interviewers' schedule of visits, the $\hat{X}_i$ and $\hat{X}$ become biased estimates. Under these conditions

$$\hat{X} = \sum_{i=1}^{6} \hat{X}_i = 6 \sum_{i=1}^{6} \sum_{j=1}^{n_i} \frac{x_{ij}}{i};$$

and

$$E\hat{X} = 6 \sum_{i=1}^{6} \sum_{j=1}^{N_i} \frac{\delta_{ij} x_{ij}}{i} \neq X$$

where δ_{ij} equals unity or zero according to whether or not the jth person in the ith stratum is at home at a moment taken arbitrarily.

If the exact probability $\bar{p}_j$ were known for all the persons and the visits are made at a random time the estimate

$$\hat{X}' = \sum_{j=1}^{n} \frac{x_j}{\bar{p}_j}$$

is an unbiased estimate for the entire population including the zero group, for

$$E\hat{X}' = \sum_{j=1}^{N_i} \frac{p_j x_j}{p_j} = \sum_{j=1}^{N_i} x_j.$$

are uncorrelated, the bias vanishes. Where the correlation in the population is linear, this estimate is unbiased. It is not, however, susceptible to proof from the sample itself that the relationship is linear within the zero interval even though extremely strong evidence is afforded by the other six groups. The problem of estimating for this group is no different than that found in a callback operation. The size of the zero group and the consequent bias will be reduced, of course, by discovering whether or not respondents are at home on more than five previous nights, just as it is reduced by making more than five callbacks.

Variance of the Sample Estimate

Our next interest is in the variance of the sample estimate, $\sigma^2_{\hat{X}}$, where the sample, according to the conditions of the experiment, depends entirely upon the number and identity of the persons at home when the interviewer calls. For the more usual case in which the entire group of N_t persons, itself constitutes a sample from a larger population, $\sigma^2_{\hat{X}}$, becomes a separate contribution to the total variance. Suppose that $K\hat{X}$ is used as an estimate of $\overline{\overline{X}}$, the aggregate value of the variate for a population out of which the N_t persons have been selected as a sample. The total variance of this estimate $K^2\bar{\sigma}^2_{\hat{X}}$ is given by:

$$\begin{aligned} K^2\bar{\sigma}^2_{\hat{X}} &= K^2E(\hat{X} - \overline{\overline{X}})^2 = K^2E[(\hat{X} - X) + (X - \overline{\overline{X}})]^2 \\ &= K^2[E(\hat{X} - X)^2 + E(X - \overline{\overline{X}})^2] = K^2(\sigma^2_{\hat{X}} + \sigma^2_X) \end{aligned}$$

where σ_X is the contribution to the variance arising out of any other sampling operations.

Our main interest, therefore, is in the contribution to the variance arising from the elimination of callbacks. By definition and by equation (10) we have,

$$\sigma^2_{\hat{X}} = E\hat{X}^2 - (E\hat{X})^2 = E\hat{X}^2 - X^2. \tag{13}$$

From equation (4) the $E\hat{X}^2$ is given by,

$$E\hat{X}^2 = E\left(6\sum_{j=1}^{n}\frac{x_j}{i}\right)^2 = 36E\sum_{j=1}^{n}\frac{x_j^2}{i^2} + 36E\sum_{j,k=1,(j\neq k)}^{n}\frac{x_jx_k}{i^2}. \tag{14}$$

The first term of (14) may be evaluated directly as in equation (5).

$$36E\sum_{j=1}^{n}\frac{x_j^2}{i^2} = 36\sum_{j=1}^{N_t}\bar{p}_jE_j\frac{x_j^2}{i^2} \tag{15}$$

in which, as before, E_j denotes the conditional expectation knowing that the j^{th} individual is interviewed. Hence,

$$E_j \frac{x_j^2}{i^2} = \sum_{i=1}^{6} \frac{x_j^2}{i^2} \frac{5!}{(i-1)!(6-i)!} \bar{p}_j^{i-1}(1-\bar{p}_j)^{6-i}. \tag{16}$$

Equations (15) and (16) readily yield

$$36E \sum_{j=1}^{n} \frac{x_j^2}{i^2} = 6 \sum_{i=1}^{6} \sum_{j=1}^{N_t} \frac{P_{ij} x_j^2}{i} \tag{17}$$

when P_{ij} is defined following equation (7) above.

The second term of (14) is the sum of all possible combinations, taken two at a time, of the weighted variate (x_j) for persons falling in the sample. The expected value of this term, therefore, is the sum of all such possible combinations for the entire population multiplied by their respective probabilities of occurence. First, we note that the probability that the j^{th} person is at home on any night in no way effects the probability for the k^{th} person, that is, x_j and x_k are independent and $Ex_jx_k = (Ex_j)(Ex_k)$. Thus, for the second term of (14), we have,

$$\begin{aligned} 36E \sum_{j,k=1, j\neq k}^{n} \frac{x_j x_k}{i^2} &= 36 \sum_{j,k=1,j\neq k}^{N_t} \left(\bar{p}_j E_j \frac{x_j}{i}\right)\left(\bar{p}_k E_k \frac{x_k}{i}\right) \\ &= \left[6 \sum_{j=1}^{N_t} \bar{p}_j E_j \frac{x_j}{i}\right]^2 \\ &\quad - \sum_{j=1}^{N_t} \left(6\bar{p}_j E_j \frac{x_j}{i}\right)^2. \end{aligned} \tag{18}$$

The first term of (18) has been shown to be equal to X^2 (equations (5) through (10), above). Similarly, it was shown that,

$$6\bar{p}_j E_j \frac{x_j}{i} = x_j(1 - \bar{q}_j^6). \tag{19}$$

These substitutions for the second term of (14), together with the right member of (17), for the first term of (14) produce the following expression for $E\hat{X}^2$:

$$E\hat{X}^2 = 6 \sum_{i=1}^{6} \sum_{j=1}^{N_t} \frac{P_{ij} x_j^2}{i} + X^2 - \sum_{j=1}^{N_t} x_j^2 (1 - \bar{q}_j^6)^2. \tag{20}$$

Substituting in (13) and simplifying, we have,

$$\sigma^2_{\hat{X}} = \sum_{j=1}^{N_t} x_j^2 \left\{ 6 \sum_{i=1}^{6} \frac{P_{ij}}{i} - (1 - \bar{q}_j^6)^2 \right\}. \tag{21}$$

In almost all sampling problems of the type for which this technique is appropriate, the variances in the population are unknown and it is necessary to substitute values from the sample to compute the sampling error of the sample estimate. In this case $\tilde{p}_k = k/6$ is the sample estimate of $\bar{p}_j$ for all persons in the k^{th} stratum and $\tilde{q}_k = 1 - \tilde{p}_k$. The sample estimate of $\sigma^2_{\hat{\bar{X}}}$ then becomes,

$$\sigma^2_{\hat{\bar{X}}} = 6 \sum_{k=1}^{6} \sum_{j=1}^{n_k} \frac{x_{kj}^2}{k} \left[6 \sum_{i=1}^{6} \frac{\hat{P}_{ij}}{i} - (1 - \tilde{q}_k^6)^2 \right] \tag{22}$$

where $\hat{P}_{ij}$ is the sample estimate of P_{ij} obtained by substituting $\tilde{p}_k$ and $\tilde{q}_k$ for $\bar{p}_j$ and $\bar{q}_j$. Since there are only six possible values of $\tilde{p}_k$ and $\tilde{q}_k$ as listed in Table I, the corresponding six values of

$$\frac{6}{k} \left[6 \sum_{i=1}^{6} \frac{\hat{P}_{ij}}{i} - (1 - \tilde{q}_k^6)^2 \right] = A_k$$

have been worked out for use in estimating the variance of any sample estimate based on this plan. They are shown in Column Three of Table II. The sum of Column Five is the sample estimate of the variance, $\hat{\sigma}^2_{\hat{\bar{X}}}$.

TABLE II

Stratum (K) (1)	$6/K$ (2)	A_k (3)	S_k* (4)	$\hat{\sigma}_{\hat{\bar{X}}_k}$ (5)
1	6	16.160	S_1	16.160S_1
2	3	6.957	S_2	6.957S_2
3	2	2.802	S_3	2.802S_3
4	1.5	1.027	S_4	1.027S_4
5	1.2	.305	S_5	.305S_5
6	1	0	S_6	—

* $S_k = \sum_{j=1}^{n_k} x^2_{kj}$.

Numerical Examples

To gain insight into the probable effects of practical applications of this plan, it may be helpful to consider a hypothetical population and the expected sample which would result. Such a population is shown in Table III. The probability density function of the distribution of $\bar{p}_j$ in this population is $3\bar{p}_j^2 d\bar{p}_j$. This function was selected primarily for convenience in computing the expected numbers at home $1, 2, \cdots, 6$ nights out of the six in question. It does, however, roughly approximate a typical distribution which might be inferred from actual answers to the questions regarding nights at home.

TABLE III

HYPOTHETICAL DISTRIBUTION OF THE POPULATION AND EXPECTED SAMPLE BY PERCENT OF INTERVIEWING HOURS PERSONS ARE AT HOME, AND THE EXPECTED NUMBER OF NIGHTS AT HOME OUT OF SIX SPECIFIED NIGHTS

	Expected No. of Nights Home out of Specified Six	0–10%	10%–20%	20%–30%	30%–40%	40%–50%	50%–60%	60%–70%	70%–80%	80%–90%	90%–100%	0%–100%	p_i
Population		1,000	7,000	19,000	37,000	61,000	91,000	127,000	169,000	217,000	271,000	1,000,000	.75
	0	630	2,486	3,278	2,751	1,690	772	247	47	4	—	11,905	.30
	1	298	2,762	6,597	8,823	8,168	5,520	2,645	794	107	2	35,716	.40
	2	63	1,335	5,658	11,982	16,673	16,673	11,982	5,658	1,335	63	71,422	.50
	3	8	356	2,647	8,817	18,309	27,227	29,411	21,989	9,202	991	119,047	.60
	4	1	56	710	3,704	11,573	25,347	41,258	49,173	37,293	9,462	178,577	.70
	5	—	5	104	842	3,933	12,753	31,365	60,051	84,658	56,289	250,000	.80
	6	—	—	6	81	564	2,708	10,092	31,288	84,401	204,193	333,333	.90
Total Not in Sample		925	5,875	14,125	23,875	33,325	40,675	44,125	41,875	32,125	13,075	250,000	
Expected Sample		75	1,125	4,875	13,125	27,675	50,325	82,875	127,125	184,875	257,925	750,000	.80
	1	50	460	1,099	1,471	1,361	920	441	132	18	—	5,953	.40
	2	21	446	1,886	3,999	5,558	5,558	3,994	1,886	445	21	23,810	.50
	3	4	178	1,324	4,409	9,199	13,614	14,706	10,995	4,601	496	59,524	.60
	4	—	37	473	2,468	7,715	16,898	27,504	32,782	24,862	6,308	119,047	.70
	5	—	4	87	702	3,278	10,627	26,138	59,042	70,548	46,907	208,333	.80
	6	—	—	6	81	564	2,708	10,092	31,288	84,401	204,193	333,333	.90

The upper half of Table III shows the relation between the actual percentage of time persons are at home and the expected numbers at home 1, 2, $\cdots$, 6 nights out of six. The marginal distribution of $\bar{p}_j$ for the entire population (first line of Table III), is obtained by integrating successively

$$3N_t \int_a^b \bar{p}_j^2 d\bar{p}_j \qquad (0 \leqq a \leqq b \leqq 1)$$

where a and b are the percentages shown in the heading. The average per cent of the time all persons were actually at home ($\bar{p}$) is given by:

$$\bar{p} = E\bar{p}_j = 3 \int_0^1 (\bar{p}_j)\bar{p}_j^2 d\bar{p}_j = 3\left[\tfrac{1}{4}\bar{p}_j\right]_0^1 = 75\%. \tag{23}$$

The product, N_T multiplied by this integral between the limits (a) and (b), will yield the expected number falling in sample, who were at home between a and b per cent of the time. When (a) and (b) are taken successively in intervals of 10%, the distribution of ($\bar{p}_j$) in the sample is obtained (line opposite "Expected Sample" in Table III). Similarly, the expected numbers not in the sample are obtained from the integral

$$3N_t \int_a^b \bar{p}_j^2(1 - \bar{p}_j)d\bar{p}_j = 3N_t\left\{\int_a^b \bar{p}_j^2 d\bar{p}_j - \int_a^b \bar{p}_j^3 d\bar{p}_j\right\} = N_t - n.$$

The expected numbers, in the population, at home 1, 2 $\cdots$, 6 nights out of six are given by:

$$\begin{aligned} EN_i &= 3N_t \int_a^b \frac{6!}{i!(6-i)!}\bar{p}_j^i(1 - \bar{p}_j)^{6-i}\bar{p}_j^2 d\bar{p}_j \\ &= 3N_t \int_a^b \bar{p}_j^2 P_{ij} d\bar{p}_j. \end{aligned} \tag{24}$$

The expected numbers, for the population, in the main body of the table are obtained by using the successive limits (a) and (b) in intervals of 10% ($i=1, 2, 3 \cdots 6$). The corresponding expected numbers in the sample are obtained from

$$\begin{aligned} En_i &= 3N_t \int_a^b \frac{5!}{(i-1)!(6-i)!}(\bar{p}_j)\bar{p}_j^{i-1}(1 - \bar{p}_j)^{6-i}\bar{p}_j^2 d\bar{p}_j \\ &= 3N_t\left(\frac{i}{6}\right)\int_a^b P_{ij}\bar{p}_j^2 d\bar{p}_j. \end{aligned} \tag{25}$$

For any integral value of i, we have,

$$En_i = \frac{i}{6} EN_i. \tag{26}$$

Clearly this relation holds regardless of the limits of integration and the expected number in the sample in any "cell" is exactly equal to $i/6$ multiplied by the corresponding expected number in the population. Therefore, by applying the weights $6/i$ to any group in the sample, we obtain an unbiased estimate of the corresponding group in the population.

An interesting check, useful in practical applications of the plan, is afforded by comparing $n/\hat{N}$ with n/N_t since each of these ratios is an estimate of the average per cent of time all persons in the population are at home. It is clear that n/N_t, the per cent of all individuals visited who were actually found at home, is an unbiased estimate, where calls have been scheduled at random. Because $E\hat{N} = N \leq N_t$, this comparison indicates the net error resulting from several sources including (1) bias due to the zero group (2) sampling error (3) response bias in reporting information on previous nights at home, etc. For the population in Table III, the bias due to zero group equals

$$\text{Bias} = \frac{750{,}000}{988{,}095} - \frac{750{,}000}{1{,}000{,}000} = .757 - .75 = .007.$$

Several other interesting comparisons can be made from this "model" population and expected sample. Consider the characteristic x_j. which, for example, is possessed by all persons in the population who are at home more than 80% of the time and by no one else. The expected value of the sample estimate $\hat{X}$ is $E\hat{X} = X = 487{,}996$ while the population value for the entire N_T persons equals 488,000. On the other hand, if the characteristic under study were possessed by all persons in the population at home less than 20% of the time and no one else, $E\hat{X} = X = 4{,}884$ while the population value for the entire N_T persons is $X_t = 8{,}000$.[11] This latter substantial bias (3,116 = 39%), of course, would also result from a sample in which five callbacks were made.This comparison points up the fact that in extreme cases where characteristics are possessed almost exclusively by persons who go out frequently,

[11] In many situations, it may be preferable to use the ratio estimate $\hat{X}' = (N_t\hat{X}/\hat{N})$ instead of $\hat{X}$. Where the primary interest is in the per cent (X_t/N_t), the estimate $(\hat{X}/\hat{N}) = (\hat{X}'/N_t)$ is often to be preferred. This estimate contains a bias, which is usully not large, arising from the random variate $\hat{N}$ appearing in the denominator. For the two examples, the estimate $\hat{X}'$ yields, respectively, 493,875 and 4,943.

extraordinary effort must be made to reduce the not-at-home bias. In such a case, this plan may be coupled with—say two callbacks to those not at home on the first call, in which questions concerning previous not-at-home performance are asked, as in the original calls. Application of similar projection procedure to these respondents will reduce the bias to about 30%. From the standpoint of sampling bias alone, therefore, in an extreme case this technique together with two callbacks will reduce the bias of 39% obtained from five callbacks, to a bias of 30%, after eliminating three expensive callbacks.

Obviously, the considerable advantage afforded by this plan in reducing bias must be modified to some extent by consideration of the contribution to sample error because of the failure to include all N cases. For the two examples just cited, this sampling error as estimated from the sample cases is .00059 and .0225 (coefficient of variation), for the cases, (respectively), in which the characteristic is possessed by persons at home more than 80% of the time and less than 20% of the time.

In contemplating any particular survey operation, it is necessary to compare the possible benefits which might result from the use of the nights-at-home questions with those of alternative plans, as for example, provision for callbacks to only a sub-sample of the persons not at home. The sub-sample, of course, also contributes to the sampling error. Unfortunately, there are few safe generalizations regarding which plan will produce the most information for the cost. In a real sense, the two plans are equivalent, for the nights-at-home questions result in sub-samples of individuals who are usually at home about one night in six, two nights in six, and so on. Moreover, the number of sample cases obtained for each of these "sub-samples" tends to correspond roughly to the numbers that would have been obtained from sub-samples of callbacks that were allocated according to optimum conditions considering the costs. That is to say, second call interviews cost more than first call interviews; and third, fourth, fifth and sixth call interviews become progressively more and more expensive so that the optimum allocation formula leads to smaller and smaller sub-samples of callbacks for each of these groups, respectively. This is because respondents found on callbacks tend to be more widely scattered requiring extra travel time and expense.

Since the addition to the variance resulting from the use of the nights-at-home question is only one contribution to the over-all variance of the estimate, its importance depends upon the efficiency of the original sample. If the original sample is both large and highly efficient

it may pay to obtain additional interviews from persons frequently away from home even at a high cost per interview. Experience indicates, however, that population samples are seldom efficient enough to warrant the extra cost. This naturally depends also upon the particular cost structure of the organization conducting the survey.

Because of the effects of clustering respondents to save travel costs, efficiently conducted population surveys are likely to have more than twice the sampling error of a widely scattered unrestricted random sample of the same size. By making this assumption, it is possible to compare the results of this plan for eliminating callbacks, Plan A, with those of an operation providing for initial calls to 10,000 persons and up to five callbacks to find the not-at-homes, Plan B. The two samples are to be drawn from the population described in Table III. The characteristic under study is possessed by about half of the population and is unrelated to the tendency to be at home. Under these conditions the most probable results of these two operations are shown in Table IV.

The total number of home visits under each plan are equal. Inasmuch as all visits under Plan A are first calls, while about 4,500 visits under Plan B are callbacks, Plan B must require more expensive field work. Since the sampling error for Plan B is larger than for Plan A, it is apparent that Plan A yields more information for the costs, under the stated assumptions of this example. If the original sample were twice as efficient as a random sample, the sampling error for Plan A would be almost exactly equal to that of Plan B, one-fourth of one per cent. Thus, Plan A would still yield more information for the cost. The perfect correlation, $\rho_{\hat{x}\hat{N}}=1$ implicit in the assumption that the characteristic under study is not related to the frequency with which persons are away from home, tends to reduce the sampling error of Plan A somewhat unrealistically. It is easy to construct hypothetical examples based on other assumptions to compare the two plans and possibly to appraise the effects of sub-sampling the original not-at-homes.

Some allowance should be made for the fact that interviewers are frequently able to find out from other members of the family when the absent member will be at home and by scheduling callbacks according to this information a higher percentage of persons may be found on subsequent visits. The primary difficulty here, however, is that callbacks necessitate travel to a neighborhood and it is seldom possible to schedule the visits to coincide with the various times that several persons to be interviewed in the neighborhood are at home. In addition, no one will be found at home in many of the households visited and it is not easy to obtain reliable information from neighbors. Nevertheless,

TABLE IV

COMPARISON OF THE PROBABLE RESULTS OF A CALLBACK OPERATION WITH THOSE OF A PLAN FOR ELIMINATING CALLBACKS

Plan A. Nights at Home Questions					Plan B: Callback Operation			
No. of nights at home out of six	Number of visits	Inter-views	Number with characteristic (x)	$\hat{X}_i$		Number of visits	Not-at homes	Inter-views
(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)
All	14,464	10,849	5,425	7,146	All calls	14,464	4,583	9,881
6	4,822	4,822	2,411	2,411	1st Call	10,000	2,500	7,500
5	3,616	3,014	1,507	1,808	2nd Call	2,500	1,000	1,500
4	2,583	1,722	861	1,292	3rd Call	1,000	500	500
3	1,722	861	431	861	4th Call	500	286	214
2	1,032	344	172	516	5th Call	286	178	108
1	517	86	43	258	6th Call	178	119	59
0	172	—	—	—				

$$\tilde{p}_A = \frac{\hat{X}}{\hat{N}} = \frac{7,146}{14,464 - 172} = 50\%$$

$$\tilde{p}_B = 50\%$$

$$\sigma_{\tilde{p}_B} = 2\sqrt{\frac{(.50)(.50)}{9,881}} = 1.006\%$$

$$\sigma_{\hat{X}}^2 = 4,443 \quad \sigma_{\hat{N}}^2 = 8,886 \quad \rho_{\hat{X}\hat{N}} = 1.00$$

$$\sigma^2_{\hat{X}/} \cong \frac{\hat{X}^2}{\hat{N}^2}\left\{\frac{\sigma_{\hat{X}}^2}{\hat{X}^2} + \frac{\sigma_N^2}{\hat{N}^2} - \frac{2\rho_{\hat{X}\hat{N}}\sigma_{\hat{X}}\sigma_{\hat{N}}}{\hat{X}\hat{N}}\right\} \cong .00000186$$

$$\sigma_{\tilde{p}_A}^2 \cong \sigma^2_{\hat{X}/\hat{N}} + 4\left[\frac{(.50)(.50)}{14,292}\right] \cong .00000186$$

$$+ .00006997 \cong .00007183$$

$$\sigma_{\tilde{p}_A} \cong .85\%.$$

to the extent that such information can be found and utilized the callback operations become more effective.

Other considerations pertinent to a decision regarding the possible use of the night-at-home questions may be mentioned briefly, as follows: (a) the efficiency of the original sample, (b) the number of callbacks permitted by the budget, (c) the time available in which to produce the final survey results, (d) the relationship between the variate under study and the tendency to be away from home, (e) the time of day during which interviews are to be conducted, (f) the particular population group under study, e.g., farmers, housewives, car owners, men over 21 years old, and so on. It should prove to be especially helpful in the further development of this plan to have reports from others concerning their experience with it, as well as any suggestions for modifications or improvements in the particular techniques described above.

APPLICATION OF LEAST SQUARES REGRESSION TO RELATIONSHIPS CONTAINING AUTO-CORRELATED ERROR TERMS*

D. COCHRANE AND G. H. ORCUTT
Department of Applied Economics, Cambridge

We point out that autocorrelated error terms require modification of the usual methods of estimation and prediction; and we present evidence showing that the error terms involved in most current formulations of economic relations are highly positively autocorrelated. In doing this we demonstrate that when estimates of autoregressive properties of error terms are based on calculated residuals there is a large bias towards randomness. We demonstrate how much efficiency may be lost by current methods of estimation and prediction; and we give a tentative method of procedure for regaining the lost efficiency.

INTRODUCTION

THREE MAJOR complications may be distinguished in the statistical measurement of relationships between economic time series:

1. The existence of simultaneous relationships between the variables.
2. The presence of auto-correlated error terms. This has been less exactly called the time-series complication.
3. The presence of errors of observation in each of the variables.

The first complication was forcefully brought to the attention of economists by Frisch [1] and Haavelmo [2]; and much work has since been done by Koopmans [3] and others, [4] in finding the structural parameters when the economic variables are described by a system of simultaneous equations. This approach is very promising but the time-series complication has been assumed away by the specification that the error terms which enter into each equation are independent in successive periods of time.

A considerable amount of work has also been devoted to problems relating to the second complication. The rather extensive literature connected with the variate difference method, conveniently summarized by Tintner [5], and also the general analysis of economic time trends

* We wish to express our thanks for the considerable assistance we have received from Richard Stone.

may be included under this heading. More directly related to the problem are the studies which examine the distribution of correlations between autocorrelated series, [6] the major proportion of which are devoted to tests for the null hypothesis. Of those papers which are concerned with the measurement of functional relationships between series, few make it clear that the significant factor in the analysis is the autocorrelation of the error term and not the autocorrelation of the time series themselves. This fact has been well expressed by Aitken, [7] but its importance seems to have escaped the attention of economists. We should also refer to a paper by Champernowne [8] which became available after this study was essentially complete. Champernowne's paper recovers much of the ground developed by Aitken and is an exceedingly useful study, carrying the problem into the field of statistical estimation and sampling theory.

The third complication arises when the assumption that the explanatory variables are measured free from error cannot be maintained, and may therefore be a problem of some importance when considering economic data. In the absence of a complete knowledge of the correlation matrix of the errors, simplifying assumptions that the errors in each of the variables are random and uncorrelated both with the systematic part of each variable and with the errors in the other variables must be made. The problems involved have received consideration in the work of Frisch [9], Koopmans [10], Tintner [11], Reiersøl [12] and Geary [13].

The objects of this paper are four-fold. First, we wish to focus the attention of economists on the fact that the presence of autocorrelated error terms requires some modification of the usual least squares method of estimation; and secondly, we wish to show that there is strong evidence in favour of the view that the error terms involved in most current formulations of economic relations are highly positively autocorrelated. In doing this we demonstrate the presence of a large bias towards randomness in estimates of the autoregressive properties of error terms which are based on calculated residuals. Third we indicate roughly how much efficiency is lost by current methods of estimation and prediction if error terms are highly autocorrelated; and finally present a tentative method of procedure.

In arriving at our conclusions we have placed considerable reliance on results obtained from a number of sampling experiments. We recognize that results arrived at by this procedure may not have the elegance or all of the utility of results obtained deductively from the same assumptions; nevertheless this method of approach is a legitimate one

and frequently makes it possible to obtain useful answers to problems which have proved stubborn to mathematical statisticians. In this connection it might be noticed that there are a large number of important questions in the field of statistics which in principle could be answered deductively but which have till the present time proved too difficult. Most of these questions could be answered by sampling experiments and it is to be hoped that, as improved calculating equipment becomes available, more attention will be given to this approach.

In order to concentrate on the problem of auto-correlated errors, we have ignored the difficulties arising from the simultaneous equations complication and the errors in the variables complication. However, it should be obvious that for the purpose of estimating structural parameters it is necessary to find a method of dealing simultaneously with all three complications, or at least some indication of their relative importances. A consideration of some aspects of the difficulties to be experienced in analysing relationships when more than one of these complications are present is contained in a following paper [14].

REGRESSION ANALYSIS WITH AUTOCORRELATED ERROR TERMS OF KNOWN AUTOREGRESSIVE PROPERTIES

It may be helpful to restate briefly the assumptions underlying the method of least squares. Suppose a single linear relationship exists between the variables $x_{1t}, x_{2t}, \ldots x_{kt}$ of the form

$$x_{1t} = a + \sum_{j=2}^{k} b_{1j} x_{jt} + u_t \tag{2.1}$$

where u_t is a random error term with constant variance, while the a and the b's are constants to be determined. Provided the $x_{2t} \ldots x_{kt}$ are independent of the random error term u_t, then the best linear unbiassed estimates of these coefficients are given by the method of least squares, best estimates meaning those estimates which have a minimum variance. This is true even if the independent variables are autocorrelated, provided we can consider them as fixed in repeated samples [15]. If in addition the error term is normally distributed then the least squares estimates are maximum likelihood estimates [16].

In many economic relationships it is an oversimplification to assume that error terms are independent in time. If we have a relationship in

which the error term is autocorrelated, it has been shown by Aitken [17] that the method of least squares still yields the best linear unbiassed estimates of the regression coefficients provided the lack of independence in the error series is taken into account. One method of overcoming this lack of independence is to make the error term random by transforming all the variables according to the autoregressive structure of the error term. Suppose we have a linear relationship given by

$$y_t = a_0 + a_1x_t + u_t \tag{2 2}$$

where u_t is generated by the Markoff scheme

$$u_t = \beta u_{t-1} + \epsilon_t \tag{2.3}$$

with random disturbances ϵ_t and a known autoregression coefficient β. We may substitute for u_t in equation (2.2) and obtain

$$y_t' = a_0' + a_1x_t' + \epsilon_t \tag{2.4}$$

where

$$y_t' = y_t - \beta y_{t-1} \quad \text{and} \tag{2.5}$$

$$x_t' = x_t - \beta x_{t-1} \tag{2.6}$$

and the application of least squares to equation (2.4) will produce best linear unbiassed estimates of the regression coefficients a_0' and a_1.[1]

It is also possible to improve on the ordinary methods of prediction when the error terms are autocorrelated. If we wish to estimate y_t from a given x_t it can be seen that equation (2.2) is not the most efficient form in which to make this estimation. A more appropriate form would be to use the relation

$$y_t = a_0' + a_1(x_t - \beta x_{t-1}) + \beta y_{t-1} \tag{2.7}$$

where a_0' and a_1 are estimated from (2.4). In a later section we shall illustrate the gain to be achieved by using this relation in problems of estimation.

In the discussion which follows it is convenient to restrict the meaning of error term to the true series of errors in a relationship, that is the series of errors which would be obtained if the *true* values of the regression parameters were applied in the relationship. To distinguish the discrepancies actually obtained from the true errors we shall call them residuals. In addition, we shall limit the word disturbance to describe the random elements in an autoregressive equation.

[1] A more complete statement of this solution is to be found in Section VI.

AUTOCORRELATION OF ERROR TERMS AND RESIDUALS OF ECONOMIC AND CONSTRUCTED RELATIONSHIPS

In this section we develop the argument that the error terms in many if not most current formulations of economic relations are highly positively autocorrelated, but it should be stressed that we are not trying to prove that this must be so in every case or that it is impossible to formulate relations in which the error terms are random. Since the autocorrelation properties of economic time series will frequently arise in this section, we should first like to refer to a study by Orcutt [18] in which it is shown that the fifty-two series used in Tinbergen's [19] model of the economic system of the United States might be considered to have been obtained by drawings from a single population of linear stochastic series having the same underlying autoregressive structure. The underlying autoregressive equation was estimated to be of the form

$$x_{t+1} = x_t + 0.3(x_t - x_{t-1}) + \epsilon_{t+1} \tag{3.1}$$

where the ϵ's are random disturbances. The high positive autocorrelation of economic time series which (3.1) implies is a feature which should not be overlooked.

Turning to the error terms, let us investigate their sources and see if there is reason to believe that the error terms also are likely to be highly positively autocorrelated. We can examine their sources under three main headings.

(1.) Systematic errors may arise from a faulty choice of the form of relationship assumed to exist between economic variables. Since the economic variables are positively autocorrelated, then in general errors of this type will be positively autocorrelated. Further the shortness of most available time series makes the statistical results meaningless if very complicated relationships are adopted, so that errors of this type are inevitable.

(2.) Error terms may arise owing to the omission of variables, both economic and non-economic, from the analysis. Important variables may be omitted either because they are not available or because their importance is not realized. Furthermore, because of the brevity of available time series, it is also frequently necessary to neglect variables which individually have but a small influence. Nevertheless, it is evident that the total influence of a number of such variables may be

very substantial and highly positively autocorrelated.[2] Now, as already indicated, there is strong eivdence in favour of believing that most economic time series are highly positively autocorrelated. Therefore, in so far as the omitted variables are economic time series, we may expect the resulting error terms to be highly positively autocorrelated.

Consider also the case of non-economic variables which are likely to influence economic behaviour but which are generally omitted. Some of those that more readily come to mind are population and its age, sex and spatial distribution, changes in cultural patterns, technological developments, exploitation and exhaustion of mineral resources including changes in soil fertility, and climatic conditions. Most of the above series have very high positive autocorrelations but even where the auto correlations are not high, as in the case of at least certain climatic conditions, it is evident that their impact on the economic system is still likely to be autocorrelated. Thus even if rainfall was really a random series, the water level in the soil, being the result of rainfall over several years, would be positively autocorrelated. We might recall in this respect the correlograms given by Wold [20] of the average yearly rainfall during the period 1867 to 1936 of four cities in or near the drainage basin of Lake Väner and the average annual water level (obtained from quarterly observations) of Lake Väner from 1867 to 1936. The correlogram of the yearly rainfall indicated a random series while that of the level of the lake indicated a positively autocorrelated series showing that, whilst the occurrence of certain meteorological factors may be random, their general influence over time may be systematic.

Now it may be reasonably argued that the economic behaviour of individuals is not completely dependent on economic variables or non-economic variables of the type we have mentioned, and that, even if an explanation incorporated in the correct manner as many as necessary of these variables, it would still not yield perfectly correct predictions.[3] No doubt this is true, and the explanatory variables needed

[2] This may be shown as follows. If we have two unrelated autocorrelated series x_t and y_t whose first autocorrelations are given by

$$\frac{\text{cov}\ (x_t, x_{t-1})}{\text{var}\ (x)} \text{ and } \frac{\text{cov}\ (y_t, y_{t-1})}{\text{var}\ (y)}$$

then if $z_t = x_t + y_t$ the first autocorrelation of z_t is given by

$$\frac{\text{cov}\ (x_t, x_{t-1}) + \text{cov}\ (y_t, y_{t-1})}{\text{var}\ (x) + \text{var}\ (y)}$$

This result may be generalised to show that the sum of any number of autocorrelated series is also autocorrelated with its first autocorrelation equal to the sum of the first lag covariances of the individual series divided by the sum of the individual variances.

[3] See for example T. Haavelmo, "The Probability Approach in Econometrics," *op. cit.* Section 11.

to complete the explanation may be of an approximately random character since they relate to such things as the physiological processes of each individual. However, it would be a mistake to infer from this that economic time series contain a significant random component, for what will obviously happen when the behaviour of a large number of individuals is averaged is that those actions of individuals which are positively correlated with the actions of others will dominate the average while those actions which are random for each individual and uncorrelated as between individuals will be averaged out.

(3.) The series of data used may not measure exactly what is required for the particular analysis. In so far as the discrepancy is one of coverage, it seems reasonable to believe that the error term involved will have much the same autoregressive properties as economic series in general. In so far as the discrepancy is more nearly what might perhaps be called a pure error of observation, it would appear more difficult to say anything about whether or not it is autocorrelated. However, on the basis of discussions with economists engaged in the construction of basic economic data, we have formed a very strong impress ion that, if an error is committed one year, it will very likely be committed again the next year and that most errors of observation are positively autocorrelated.

Let us now see whether our theory is plausible by making a brief examination of the autocorrelations of the residuals obtained in several econometric studies. These are two papers by Lawrence R. Klein, "The Use of Econometric Models as a Guide to Economic Policy," [21] and "Economic Fluctuations in the United States 1921–1941" [22]; a paper by M. A. Girshick and Trygve Haavelmo [23] and a paper by Richard Stone [24]. The measure of autocorrelation used is the ratio of the mean square successive difference to the variance of the residuals. This ratio is generally denoted by δ^2/s^2 [25] where δ^2 and s^2 are defined by

$$\delta^2 = \frac{1}{N-1}\sum_{t=1}^{N-1}(x_{t+1} - x_t)^2, \tag{3.2}$$

$$s^2 = \frac{1}{N}\sum_{t=1}^{N}(x_t - \bar{x})^2, \tag{3.3}$$

where $$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t.$$

This ratio has been calculated by Klein for the residuals in his two

papers and we have computed the ratios for the residuals in the other two papers [26]. Two ratios in each of Klein's papers have been omitted as they refer to first differences of the economic series and are not comparable for our purposes. It should be mentioned that the residuals given in Klein's paper in *Econometrica* and the residuals given by Girshick and Haavelmo were calculated by the reduced-form method which presupposes that it is possible to solve for each of a number of jointly dependent variables in terms of exogenous variables and random error terms and these random error terms are simply linear combinations of the error terms given in the original system of equations [27]. The residuals obtained from Klein's mimoegraphed paper and from Stone's paper were calculated by ordinary least squares method of regression. The total number of series considered is 43 and Table I shows them classified according to source and number of parameters used in each equation. The individual values of δ^2/s^2 are illustrated on the scatter diagrams of Figures I–IV.

TABLE I

SUMMARY OF VALUES OF δ^2/s^2 OBTAINED FOR VARIOUS RESIDUALS

Source of residuals	Number of years	Number of parameters				Total
		3	4	5	6	
Klein—Econometrica	22	2	7	2	1	12
Klein—Mimeographed study	20	1	7	1	—	9
Girshick and Haavelmo	20	2	2	1	—	5
Stone	19	4	6	6	1	17
Total		9	22	10	2	43
$P(\delta^2/s^2 < 1.24) = 0.025$		7	5	4	—	16
$P(\delta^2/s^2 < 1.37) = 0.05$		8	10	4	—	22

The probability distribution of δ^2/s^2 for a random series has been tabulated [28] for various N, where N is the number of items. This distribution is symmetrical around $2N/N-1$ so that for $N=20$ the expected value of δ^2/s^2 for a random series is 2.11. This is the horizontal dotted line shown on the diagrams. In view of the high positive autocorrelation of economic time series and the reasons given for expecting error terms to be autocorrelated, there seems little chance of obtaining a value of δ^2/s^2 around the upper tail of the distribution and, since we wish to minimize the risk of failing to reject a value of δ^2/s^2 as coming from a random population, the appropriate test would seem to involve the use of the value of δ^2/s^2 corresponding to the 5 per cent significant level, from the lower tail only. Since all our series are of approximately

AUTOCORRELATION OF RESIDUALS

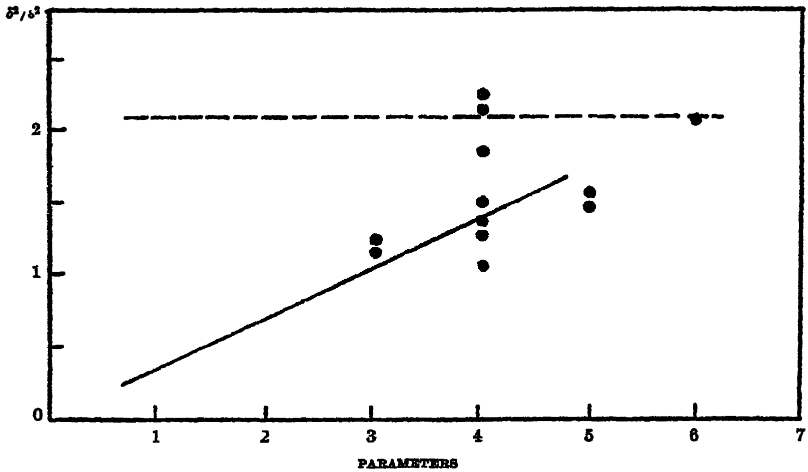


FIGURE I KLEIN-ECONOMETRICA.

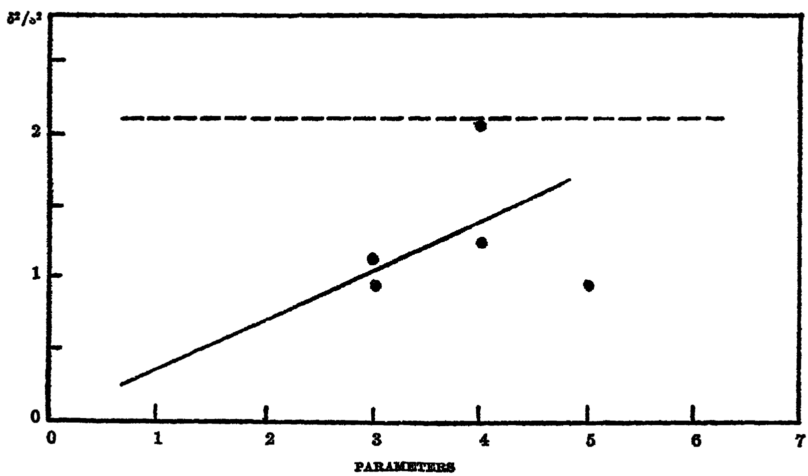


FIGURE III. GIRSHICK AND HAAVELMO.

ESTIMATED BY VARIOUS STATISTICIANS

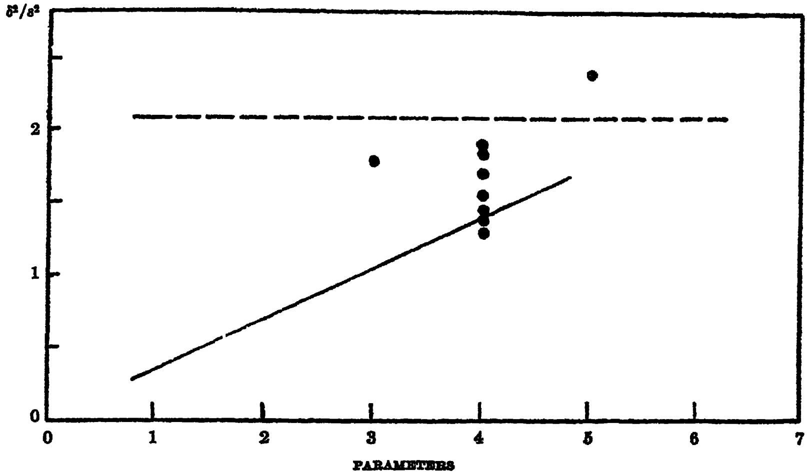


FIGURE II. KLEIN-MIMEOGRAPHED STUDY.

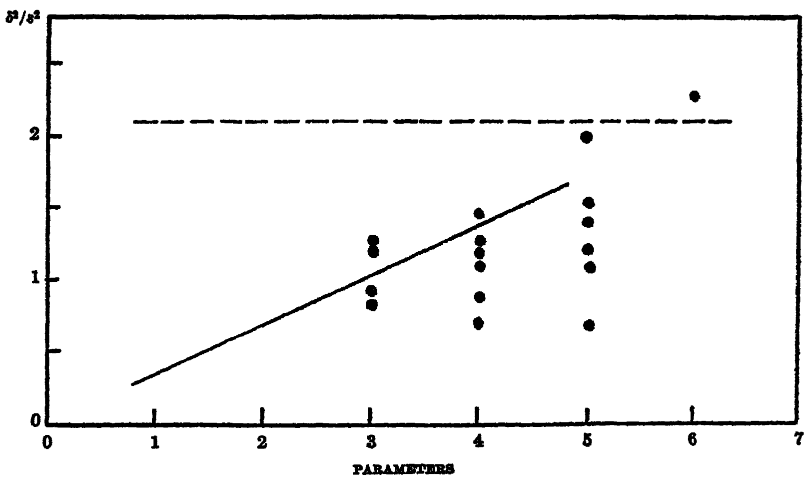


FIGURE IV. STONE.

the same length, this value is 1.37 for $N=20$. The value of $\delta^2/s^2(=1.24)$ corresponding to the 5 per cent significance level which includes both tails has also been added.[4] Out of the 43 series, 16 are significantly different from a random series at the $2\frac{1}{2}$ per cent level, while 22 are significant at the 5 per cent level. These results indicate that in many cases the assumption of random error terms is not a very good approximation to the truth.

The sloping lines on Figures I–IV correspond to the average of twenty estimates of δ^2/s^2 obtained from constructed relationships, described in subsequent paragraphs, in which the error terms were first summations of random series. It would seem more reasonable to consider that the values of δ^2/s^2 are distributed around a line of this nature rather than around the horizontal random line. This suggestion is supported by the decreasing proportion of residuals which are significantly different from random series as the number of parameters in the relationship increases. From Table I it can be seen that the proportions which are significantly different from random are 8/9, 10/22, 4/10 and 0/2 for 3, 4, 5 and 6 parameters respectively.

Construction of an experimental model. The examination of the residuals obtained from actual economic relationships fails to reject the hypothesis that error terms are highly positively autocorrelated in a number of economic relationships. Little is known about the behaviour of relationships possessing autocorrelated error terms, so it was decided to construct several relationships of this type from artificial series and observe the results of applying least squares regression. The general form of the relationship adopted was—

$$X_1 = k + b_{12.3t}X_2 + b_{13.2t}X_3 + b_{1t.23}\,t + u \qquad (3.4)$$

where X_2, X_3 and u were independently constructed series all possessing the same autoregressive structure, t represented a linear time trend and the true values of the constants were $k=0$, $b_{12.3t}=2$, $b_{13.2t}=1$ and $b_{1t.23}=0$. Thus the actual equation used for the construction was—

$$X_1 = 2X_2 + X_3 + u. \qquad (3.5)$$

Five sets of relationships of this form were constructed with different autoregressive structures, each set containing 20 equations. The series used were generated according to the following formulae:—

[4] Klein has taken the 5 per cent level of significance to include both tails of the distribution (*Econometrica, op. cit.*, p. 114).

$$
(3.6)\qquad
\begin{aligned}
&A.\quad x_{t+1} = x_t + 0.3(x_t - x_{t-1}) + \epsilon_{t+1}\\
&B.\quad x_{t+1} = x_t + \epsilon_{t+1}\\
&C.\quad x_{t+1} = 0.3x_t + \epsilon_{t+1}\\
&D.\quad x_{t+1} = \epsilon_{t+1}\\
&E.\quad x_{t+1} = \epsilon_{t+1} - \epsilon_t
\end{aligned}
$$

where the ϵ's denote series of random disturbances. Instead of stating the precise form of the autoregressive equation each time a series is referred to we shall use the letters A, B, C, D and E as a convenient notation.

The random elements were obtained from Tables of Random Sample ing Numbers [29]. Two figure numbers were extracted, ignoring the number 00, so that they ranged from 1 to 99. The number 50 was then subtracted throughout so that we possessed a rectangular distribution ranging from $+49$ to -49 with a true mean of zero. We then formed 60 independent series of these random elements, each one 20 items in length, omitting a few numbers between each series so that we could later extrapolate for forecasting. The application of these series in-groups of three to the relation (3.5) gave us the 20 equations of set D. The other transformations were then formed from this basic set. For example, the set of first summations, series B, was formed by making the first-term of each series zero and summing progressively over each item of the random set. Simplifications of the calculations involved were made by using the fact that C is the first difference of A, while B, D and E are respectively the first summation of a random series, a random series, and the first difference of a random series. It can be easily seen that there were 21 items in series A and B, 20 in C and D and 19 in E. They are therefore analogous in length to most available economic time series.

In each set a regression analysis was carried out with one explanatory variable (in this case the error term became (X_3+u)), in several of the sets the analysis was extended to two explanatory variables and in the case of set B to three explanatory variables. In addition, the statistic δ^2/s^2 was calculated for the actual error terms and for the residuals. A complete summary of these calculations is contained in Table II.

Bias introduced in estimating the autocorrelations of residuals. Given a set of equations in which the explanatory variables and the error terms possess the same autoregressive structure, can we say anything

about the way in which the autocorrelations of the residuals vary as the number of explanatory variables is increased? Figure V presents this information with each set labelled according to its autoregressive structure. The number of parameters includes the constant term so that we have one parameter when only the mean is estimated. Straight lines have been fitted visually to the points for each set using as additional points, except for D, the true values of δ^2/s^2 which are zero for A and B, 1.4 for C and 3.0 for E. The sets A, B and C show a marked bias upwards as the number of parameters is increased. It is not expected that this linearity would continue indefinitely but would flatten out as more than four parameters are used and approach nearer and nearer to the value of δ^2/s^2 expected for a random series. The random set D merely shows a distribution around the horizontal straight line and when we pass to the series of first differences of random numbers E there is only very slight evidence of a downward movement in the values of δ^2/s^2 with increasing parameters.[5]

Another way of illustrating the bias in the estimated autocorrelations of the residuals as more variables are introduced is to apply our previous test of significance to the individual values of δ^2/s^2 obtained in set B. This has been done in Table III. As the number of parameters increases the proportion of residuals which yield a value of δ^2/s^2 significantly different from that expected for a random series at the 5 per cent level grows smaller; from 19/20 when only the mean is estimated to only 10/20 when four parameters are used. This is a similar result to that found for the residuals of actual economic relationships.

TABLE III

SIGNIFICANCE TESTS APPLIED TO RESIDUALS OF SET B

Explanation	Number of Parameters	Number different from random at significance levels of		Total number os residuals
		$2\frac{1}{2}$ per cent	5 per cent	
Actual error term	1	19	20	20
	1	19	19	20
One explanatory variable	2	17	18	20
One explanatory variable+time	3	13	15	20
Two explanatory variables	3	11	14	20
Two explanatory variables+time	4	7	10	20

The amount of variance to be explained in an economic time series can be regarded as composed of two parts, the first due to the smooth movements of the autoregressive structure of the series and the second

[5] The first autocorrelation of the first differences of a random series is $r_1 = -0.5$ or $\delta^2/s^2 = 3.0$.

due to the random disturbances. What is important for a real explanation is that a proportion of the variance due to the disturbances should be explained as well as that due to the general movement of the series. Now quite high correlations between autocorrelated series may be obtained purely by chance[6] and when this happens what is largely explained is the variance due to the regular movements through time. The residuals of such a relationship will be essentially the year-to-year fluctuations and of a more random character than the original series. This can be illustrated by comparing the two cases in set B in which two explanatory variables are used, one of which includes a linear time trend and the other two real variables. From equations (4) and (5) in Table II it can be seen that, while the inclusion of time adds an amount of 0.026 less to the explanation of the variance of the dependent variable than the inclusion of the second explanatory variable, the average value of δ^2/s^2 for the residuals is 0.023 greater. These are the two points which are close together in Figure V for three parameters and it can be seen that the addition of a linear time trend in the explanation produced approximately the same bias as the inclusion of a real explanatory variable. This is confirmed by the average value of δ^2/s^2 obtained when X_2, X_3 and t are the explanatory variables.

Since the inclusion of the bogus variable time had about the same effect in biasing the residuals towards randomness as the inclusion of real explanatory variables, we were curious about the effect of including other types of non-related series in the explanation. We therefore correlated two unrelated series, X_2 and X_3, of set B. With X_2 as the dependent variable it was found that the average amount of the variance explained was 0.32, while the mean value of δ^2/s^2 for the residuals was 0.74. This latter value is slightly higher than that obtained for equation (3) of Table II where the average explained variance is 0.64 with a mean value of $\delta^2/s^2=0.69$ for the autocorrelation of the residuals. This suggests that if error terms are autocorrelated then it would frequently be a mistake to attempt to justify the statistical requirements of randomness by adding more explanatory variables or by experimenting with different combinations of the variables. Owing to the shortness of economic time series, high accidental correlations may be obtained between the variables added and the error term due to their autoregressive structures and since the residuals obtained from the least squares method of regression are orthogonal to the explanatory variables they will tend to be biassed towards a random series.

[6] See G. U. Yule *op cit.* and Orcutt and James *op. cit.*

TABLE II

SUMMARY OF STATISTICS CALCULATED FOR FIVE TRANSFORMATIONS INVOLVING RELATIONSHIPS IN WHICH THE EXPLANATORY VARIABLES AND ERROR TERM POSSESS THE SAME AUTOREGRESSIVE STRUCTURE

Equation No.	Generating Properties of		Values of δ^2/s^2				Regression Parameters								Correlation Coefficient		
			Actual error series		Estimated residuals		Constant term k		Regression Coefficients						Calculated Value		True value for infinite series
									b_{12}		b_{13}		b_{14}				
	Explanatory Variable	Error Term	Mean	Variance	Mean	Variance	Mean	Standard deviation	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	
1	A	A	0.310	0.128	0.490	0.165	16.832	155.89	2.053	0.927	—	—	—	—	0.754	0.067	0.817
2	A	A	0.310	0.128	0.790	0.153	− 5.560	169.15	2.290	0.766	—	—	−1.480	339.60	0.926	0.006	0.817
3	B	B	0.450	0.223	0.685	0.279	− 7.663	105.18	2.079	0.755	—	—	—	—	0.769	0.049	0.817
4	B	B	0.450	0.223	1.081	0.283	−16.809	112.95	2.260	0.517	—	—	−1.005	168.91	0.917	0.004	0.817
5	B	B	0.309	0.071	1.058	0.320	2.527	63.59	2.160	0.335	0.930	0.478	—	—	0.927	0.012	0.913
6	B	B	0.309	0.071	1.396	0.311	−10.074	86.35	2.174	0.220	1.062	0.157	−1.202	101.30	0.970	0.0005	0.913
7	B	B	N.C.	—	N.C.	—	N.C.	—	—	—	1.242	1.919	—	—	0.440	0.181	0.408
8	C	C	1.494	0.260	1.559	0.258	− 1.260	13.484	2.185	0.175	—	—	—	—	0.839	0.005	0.817
9	D	D	1.982	0.264	2.007	0.264	− 1.069	9.84	2.140	0.111	—	—	—	—	0.838	0.0056	0.817
10	D	D	2.138	0.228	2.153	0.171	0.102	7.46	2.120	0.075	0.973	0.051	—	—	0.930	0.0004	0.913
11	D	D	N.C.	—	N.C.	—	N.C.	—	—	—	0.963	0.305	—	—	0.381	0.044	0.408
12	E	E	2.996	0.115	3.011	0.147	0.303	2.35	2.075	0.127	—	—	—	—	0.839	0.0079	0.817
							(0)		(2.075)	(0.128)	—	—	—	—	(0.844)	(0.008)	
13	E	E	3.047	0.070	2.934	0.064	0.072	1.72	2.060	0.088	0.898	0.100	—	—	0.929	0.001	0.913
							(0)	—	(2.060)	(0.088)	(0.898)	(0.100)	—	—	(0.929)	(0.001)	
14	E	E	N.C.	—	N.C.	—	N.C.	—	—	—	0.820	0.403	—	—	0.341	0.062	0.408

* Figures in parentheses were calculated assuming a mean of zero.
N.C. indicates that certain statistics were not calculated.

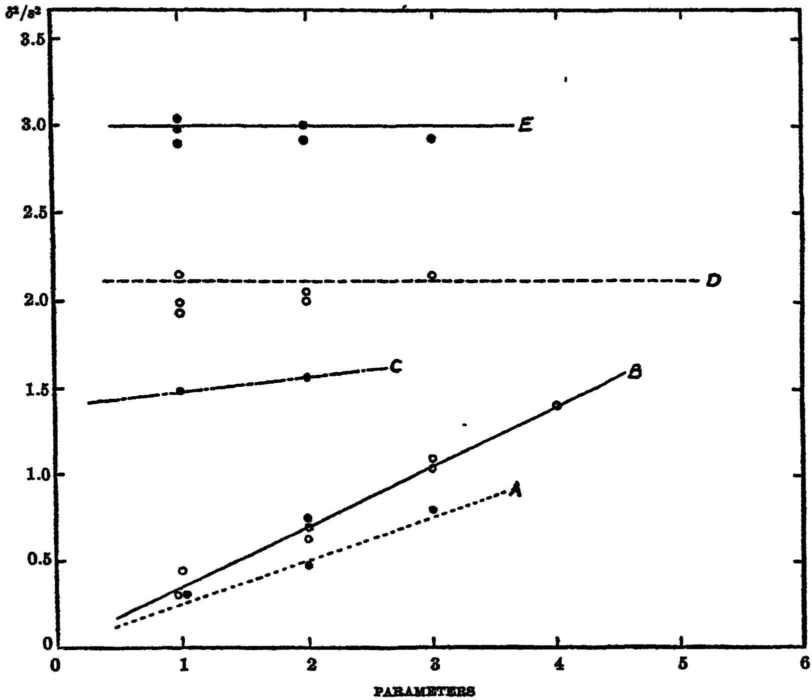


FIGURE V. AUTOCORRELATION OF RESIDUALS OBTAINED FROM CONSTRUCTED RELATIONSHIPS

ESTIMATION OF REGRESSION COEFFICIENTS AND PREDICTION BY LEAST SQUARES FOR RELATIONSHIPS CONTAINING AUTOCORRELATED ERROR TERMS

Our objectives in this section are to show that the usual application of the method of least squares to relationships containing highly positively autocorrelated error terms results in an extremely inefficient use of data and that it is only necessary to apply a transformation which will make the error term approximately random in order to regain most of this efficiency.

The complete information is contained in Table II but in order to illustrate the position more clearly we have set out some of the more relevant calculations in Tables IV and V.

TABLE IV

VARIANCES OF REGRESSION PARAMETERS UNDER DIFFERENT TRANSFORMATIONS USING ONE EXPLANATORY VARIABLE

Generating properties of		Values of $\hat{s}^2/s^2$ for		Variance of	
Explanatory variable	Error term	Error term	Residuals	Correlation coefficient	Regression coefficient
A	A	0.31	0.49	0.067	0.927
B	B	0.45	0.69	0.049	0.755
C	C	1.49	1.56	0.005	0.175
D	D	1.98	2.00	0.0056	0.111
E	E	3.00	3.01	0.008	0.127

The decline in the variances of both the correlation coefficient and the regression coefficient as the error term becomes random is very marked. In the case of one explanatory variable the variance of the correlation coefficient when the error term is of form *A* is approximately 11 times the variance when the error term is random, while the ratio of the corresponding variances of the regression coefficient is approximately 9 to 1. As we introduce more determining variables into the explanation, we can see from Table V that the variances of the regression coefficients decrease until in the limiting case all the variation in the variable to be determined is explained and there is a complete set. This limiting case is of course very rarely approached in practice and if we consider the set *B*, where for three explanatory variables the mean multiple correlation coefficient is as high as 0.97 (see Table II, equation 6), we find the variances of the regression coefficients are 0.22 and 0.16 for b_{12} and b_{13} respectively, which from Table V can be seen to be three times the variances of the regression coefficients calculated in the random transformation even though the mean multiple correlation coefficient in this form is only 0.93.

TABLE V

VARIANCES OF REGRESSION PARAMETERS UNDER DIFFERENT TRANSFORMATIONS USING TWO EXPLANATORY VARIABLES

Generating properties		Values of $\hat{s}^2/s^2$		Variance of		
Explanatory variable	Error term	Error terms	Residuals	Multiple correlation coefficient	Regression coefficients	
					b_{12}	b_{13}
B	B	0.31	1.06	0.012	0.34	0.48
D	D	2.14	2.15	0.0004	0.08	0.05
E	E	3.05	2.93	0.001	0.09	0.10

In Table IV we can see that fluctuations in the variances of the regression parameters are very small for reasonably large movements of δ^2/s^2 around the random value, given by the results for C, D and E. The true values of the autocorrelation coefficients of the error terms vary from $r_1=0.3$ to $r_1=-0.5$ in these cases. This relative stability of the variances indicates that a transformation which makes the error term approximately random will have regained most of the improvement in the efficiency possible. Similar results would also appear to be true for the case of two explanatory variables.

In our model there is no real trend, yet the introduction of a linear trend to sets A and B improves their explanation and reduces the variance of the regression coefficients. This would seem to be due to the fact already considered that the trend factor reduces the amount of autocorrelation in the residuals and can be regarded as one method of transforming the error term. In these circumstances the introduction of a polynomial trend may be a useful device in obtaining more accurate results, but it is difficult to attach an economic meaning to the coefficients of time.

In order to obtain some idea of the accuracy of estimation of regression parameters under other possible types of relationships and to illustrate once more the importance of having the error term random, we constructed from the series already calculated two sets of relationships in which the autoregressive structure of the explanatory variable and the error term were different. The form of the relationship was—

$$X_1 = k + b_{12}X_2 + v \tag{4.1}$$

where X_2 was of form A in both sets and v adopted first form B and second form D. The true values of the constants were $k=0$ and $b_{12}=2$ while the error term was taken from our previous sets with $v=X_3+u$. The first differences of each set were calculated and then a further correction was made to randomize the explanatory variable. This latter process produced error terms generated by the following formulae—

$$\text{(4.2)} \qquad \begin{aligned} &F. \quad x_{t+1} = \epsilon_{t+1} - 0.3\epsilon_t \\ &G. \quad x_{t+1} = (\epsilon_{t+1} - \epsilon_t) - 0.3(\epsilon_t - \epsilon_{t-1}) \end{aligned}$$

where the ϵ's denote a series of random disturbances. The results of the calculations are set out in Table VI and the values of δ^2/s^2 provide additional points for Figure V. In each set it can be seen that a considerable gain is to be obtained in the efficiency of the estimates of the

correlation coefficients and regression coefficients when the error term is random. If error terms are really random as postulated by many economists, there is nothing to be gained from making any transformation, even though the original series possess high positive autocorrelation. It can also be seen from the mean values of the regression coefficients of Tables II and VI that the least squares estimates are not biassed when the error term is autocorrelated even though they are not the best estimates.

Tests of Significance. It is well recognized that the ordinary test of significance for the null hypothesis can be applied to the correlation between two series provided one of them is random.[7] This can be seen to be equivalent to making the error term random in the special case of a zero regression coefficient. To apply confidence limits it is necessary that the dependent variable is distributed normally and randomly around a linear function of the explanatory variable. This is true even if the explanatory variable is not random.[8] If economic time series possess the properties which we are suggesting, then the transformation to make the error terms random will put them in a form in which it will be possible to apply confidence limits and test the significance of regression parameters in the ordinary way.

Prediction. Prediction is one of the primary reasons for undertaking statistical analysis. In Table VII we present some material derived from our constructed relations which emphasizes the huge improvement that it is possible to make if one is dealing with a formulation involving error terms which are a first summation of random elements. This table also indicates how misleading the variance of the residuals may be in such a case.

The fact that the items in column IV are smaller than those in column V is, of course, to be expected, since the regression parameters have been chosen to minimize the mean square of the residuals and the true errors are those obtained by use of the true values of the regression parameters. In the cases of random error terms, rows 2 and 5, this downward bias is small and could, if desired, be easily compensated by taking account of the number of parameters fitted. In the cases of error terms which are the first summation of random numbers, the downward bias is exceedingly large for series of this length and should

---

[7] See M. S. Bartlett, "Some Aspects of the Time-Correlation Problem in regard to Tests of Significance," *Journal of the Royal Statistical Society*, Vol. 98, 1935, pp. 536–543.

[8] See R. A. Fisher, "The Goodness of Fit of Regression Formulae and the Distribution of Regression Coefficients," *Journal of the Royal Statistical Society*, Vol. 85, 1922, pp. 597–612, and H. Cramer, "Mathematical Methods of Statistics," *op. cit.*, pp. 548–555.

emphasize the caution needed in interpreting standard errors of estimate if the error terms are likely to be highly positively autocorrelated. Column VI gives the variance of the errors of prediction one item beyond the parts of the series utilized for estimating the regression parameters. That is, each of the series involved in each set of twenty equations was extended one item and the dependent variable then predicted with a knowledge of the regression coefficients previously calculated. Column VI again illustrates in a rather simple way how misleading the variance of residuals may be when the error terms are autocorrelated, as in rows 1, 3 and 4. It should of course be realized that the much smaller variances obtained in rows 2 and 5 are due both to the fact that better estimates of the regression parameters have been obtained and used in these cases and also that the prediction formula makes use of the fact that the errors involved in rows 1, 3 and 4 are the first summation of random numbers. Thus, whereas in row 1 the estimating formula was

$$X_{1,n+1} = a_1 + b_{12}X_{2,n+1}, \tag{4.3}$$

in row 2, the estimating formula was

$$X_{1,n+1} = a_1' + b_{12}'(X_{2,n+1} - X_{2n}) + X_{1n}. \tag{4.4}$$

The errors involved in the prediction formula (4.4) are therefore random in time whereas those in (4.3) are first summations of random terms.

TABLE VII

A COMPARISON OF THE VARIANCES OF RESIDUALS, TRUE ERRORS AND PREDICTIONS OBTAINED FROM SEVERAL TRANSFORMATIONS OF THE CONSTRUCTED RELATIONS

No.	Generating properties of: Explanatory variable I	Generating properties of: Error term II	Number of explanatory variables III	Mean variance of residuals IV	Mean variance of true errors V	Variance of errors of predictions one item beyond sample VI
1	B	B	1	5142	7725	7479
2	D	D	1	1375	1466	933
3	B	B	2+time	784	4386	7127
4	B	B	2	1690	4386	3991
5	D	D	2	634	749	774

A TENTATIVE METHOD OF PROCEDURE

Having recognized that the error terms implicit in many current formulations of economic relations are highly positively autocorrelated, and also having recognized the importance of carrying out estimation and prediction by means of relations involving random error terms, how shall we proceed when faced with a practical situation? One way of evading this problem would be to change some of the variables, add additional variables, or modify the form of the relation until a relationship involving what appear to be random error terms is found. However, while this may possibly be a satisfactory way out in some cases, it obviously does not help much if by some means or other one has arrived at what is believed to be the most reasonable choice of variables and form of relation. This choice of variables and form of relation usually does not involve any specification of whether or not the errors are autocorrelated and what is required is the best method of estimating the parameters and various standard errors of the chosen relation, and not some other relation. In this situation the objective, of course, is to make an autoregressive transformation of the dependent and independent variables such that the error term becomes random. If the autoregressive properties of the error term were known, then it would simply be a matter of making the indicated autoregressive transformation as illustrated in section 2. The real problem arises when the autoregressive properties of the error term are not known but must be estimated. Except for the fact, which our experiments demonstrate, that nearly optimum results can be achieved if the error term is only a rough approximation to a random series, solution of the problem would seem rather hopeless for series of only twenty items.

One fairly obvious procedure, which we are inclined to rule out because of the large biases demonstrated in section 3, would be the following iterative process. First estimate the desired regression coefficients by ordinary least squares and obtain the resulting series of residuals. Then estimate from those residuals by least squares the autoregressive parameters of a one or two lag difference equation. Use these autoregressive parameters to make an autoregressive transformation of the observed series aimed at randomizing the error term, and re-estimate the desired regression coefficients. Put these revised estimates back in the original equation, obtain the resulting series of residuals and estimate their autoregressive parameters. Use these to make a new autoregressive transformation of the original series and so on until estimates of the desired regression coefficients are obtained which

are consistent with estimates of the autoregressive parameters of the residuals in the sense that no further adjustments are necessary. Since it is only necessary to make the error term approximately random it is unlikely that much would be gained by carrying the above process more than one or two rounds. The real difficulty with this procedure is that the series of residuals will, as shown in section 3, be strongly biassed towards randomness and therefore the autoregressive transformation based in the above way on the residuals may not in fact go far enough in randomizing the error term.

An alternative procedure which appears more promising to us is that of selecting an autoregressive transformation of the series involved such that the autocorrelations of the series of residuals are approximately equal to the expected values of autocorrelations of random series of the same length. We have not worked out an efficient procedure for doing this; but, if one is willing to approximate the autoregressive properties of the error term by a one or even two lag linear difference equation, it is fairly easy after one or two trials to choose an autoregressive transformation which will result in residuals that are sufficiently random. Furthermore, if our evidence that many error terms appear to be approximately first summations of random term is accepted, then the obvious procedure is to work with first differences of the series used. Thus, given a relation between ordinary economic variables

$$X_{1t} = a_1 + b_{12}X_{2t} + b_{13}X_{3t} \tag{5.1}$$

we suggest as a first approximation estimation and prediction in the form

$$(X_{1t} - X_{1,t-1}) = b_{12}(X_{2t} - X_{2,t-1}) + b_{13}(X_{3t} - X_{3,t-1}). \tag{5.2}$$

If (5.1) had contained a linear trend then (5.2) would have contained a constant term. The residuals from (5.2) can be obtained and tested for randomness.

If we prove to be right about the nature of most error terms in current formulations of economic relations, then the residuals of the first difference transformation will turn out to be sufficiently random and no further steps will be necessary. If the residuals in this form do not turn out to be sufficiently random, then a new transformation can be devised on the basis of their autocorrelations. The main advantages of this procedure are, first, that in many cases it will result immediately in the correct transformation and, secondly, that when it does not it will usually result in residuals that are not highly positively

autocorrelated and thereby reduce the amount of bias towards randomness which is present in this case. This will be a help in devising successive autoregressive transformations.

On the basis of this study Richard Stone[9] has recalculated a number of demand studies for the United Kingdom 1920–38. The general results will be published by Stone, but he has kindly made available to us the material presented in Table VIII. We present this material as further evidence that in many cases the use of first differences does result in essentially random series. It also seems reassuring, in so far as Stone's work is concerned, and rather remarkable, that in most cases the multiple correlations for the relations in first difference form remained very high.

TABLE VIII

VALUES OF δ^2/s^2 FOR A NUMBER OF DEMAND STUDIES FOR THE UNITED KINGDOM 1920–38

Commodity	Number of parameters	Values of δ^2/s^2 for residuals		Adjusted multiple correlation coefficient	
		Original data	First differences	Original data	First differences
Beer	3	1.28	1.86	0.989	0.962
	4	1.13	2.01	0.989	0.977
	4+time	1.23		0.993	
Spirits	3+time	1.26	2.63	0.992	0.875
Telegrams	3	1.24	1.61	0.985	0.967
	4+time	1.10	1.65	0.987	0.966
Imported wine	4	1.49	1.84	0.893	0.754
Communication services	3+time	0.71	2.05	0.996	0.834
	4+time	0.70	2.11	0.996	0.822
Lard	3+time	0.90	2.06	0.838	0.864
Margarine	4	1.26	1.80	0.959	0.748
	4+time	2.02		0.969	
	5+time	2.31	2.31	0.976	0.756
Mean value of δ^2/s^2		1.28	1.99		

APPENDIX TO SECTION II

It is of interest to compare the simple solution presented in section II with the general solution given by Aitken [30]. We shall not repeat

[9] These studies were originally given in his paper on "Analysis of Demand," *op. cit.*, but the recalculations were made on the basis of revised estimates of the data.

his elegant and rigorous proofs but shall merely illustrate his approach and deduce the special case where the error series follows a simple Markoff scheme. For this it is necessary to follow his generality of notation and employ matrices and vectors, using P' and y' to denote the matrix or vector obtained by transposing P or y and P^{-1} as the inverse matrix of P.

Consider first the simple case of least squares with non-autocorrelated errors. Let the approximate representation of the column vector of data

$$y = \{y_1 y_2 \cdots y_n\} \tag{6.1}$$

by the column vector

$$z = \{z_1 z_2 \cdots z_n\} \tag{6.2}$$

be linear in terms of a set of $(k+1)$ prescribed functions

$$1,\ x_{1t},\ x_{2t},\ \cdots,\ x_{kt} \qquad (t = 1, \cdots n). \tag{6.3}$$

Let P denote the matrix of these functional values so that the ith row of P is the row vector

$$[1,\ x_{1i},\ x_{2i} \cdots x_{ki}]. \tag{6.4}$$

Then P is of order $n\times(k+1)$ and with the restriction of linear independence over the n values $x_{i1}, \cdots, x_{in}$, it is of rank $(k+1)$. Let a denote a column vector of $(k+1)$ coefficients

$$a = \{a_0 a_1 a_2 \cdots a_k\}. \tag{6.5}$$

Then the set of values z_i is the vector

$$z = Pa. \tag{6.6}$$

If the data y are independent then the principle of least squares minimizes the sum of the squared residuals. This is the vector product

$$s^2 = (y - Pa)'(y - Pa) \tag{6.7}$$

and for the minimal conditions $\partial s^2/\partial a = 0$ we obtain the set of normal equations

$$P'Pa = P'y. \tag{6.8}$$

Having established this general result for least squares, Aitken extends the argument to the case of autocorrelated errors. If the set of errors be arranged according to their variances and covariances by the elements of a symmetric matrix U of order $n\times n$, then the least squares

estimates are obtained by minimizing

$$(6.9)\qquad (y - Pa)'U^{-1}(y - Pa).$$

Differentiating in the manner above, we obtain the set of more general normal equations

$$(6.10)\qquad P'U^{-1}Pa = P'U^{-1}y.$$

Let us now apply these general results to a simple specific example. Suppose we have a linear relation

$$(6.11)\qquad Y_t = \alpha_0 + \alpha_1 X_t + u_t \qquad (t = 1 \cdots n)$$

where u_t is defined by the simple Markoff process

$$(6.12)\qquad u_t = \beta u_{t-1} + \epsilon_t \qquad (\beta < 1)$$

where β is a known constant of ϵ_t a random disturbance. Our variance, covariance matrix of error may be defined by the symmetric matrix of order $n \times n$ where we have assumed unit variance of ϵ_t for simplicity, although the final result would not be altered if we did not.

$$(6.13)\qquad U = \frac{1}{1-\beta^2}\begin{bmatrix} 1 & \beta & \beta^2 & \cdot & \beta^n \\ \beta & 1 & \beta & \cdot & \beta^{n-1} \\ \beta^2 & \beta & 1 & \cdot & \beta^{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \beta^n & \beta^{n-1} & \beta^{n-2} & \cdot & 1 \end{bmatrix}$$

from which we obtain the symmetric inverse matrix

$$(6.14)\qquad U^{-1} = \begin{bmatrix} 1 & -\beta & 0 & \cdot\cdot & & 0 \\ -\beta & 1+\beta^2 & -\beta & \cdot\cdot & & 0 \\ 0 & -\beta & 1+\beta^2 & \cdot\cdot & & 0 \\ \cdot & \cdot & \cdot & \cdot & & \\ & & & \cdot & 1+\beta^2 & -\beta \\ 0 & 0 & 0 & \cdot & -\beta & \end{bmatrix}$$

The matrix P is of order $n \times 2$ where the ith row is

$$(6.15)\qquad [1 \; X_i]$$

while the vector of coefficients becomes the column vector

$$(6.16)\qquad \begin{bmatrix} \alpha_0 \end{bmatrix}$$

Applying these components to the general normal equations (6.10) and expanding we obtain the estimate of α_1 as

$$\hat{\alpha}_1 = \frac{\sum_1^n x_t y_t - \beta \sum_2^n x_t y_{t-1} - \beta \sum_2^n x_{t-1} y_t + \beta^2 \sum_3^n x_{t-1} y_{t-1}}{\sum_1^n x_t^2 - 2\beta \sum_2^n x_t x_{t-1} + \beta^2 \sum_3^n x_{t-1}^2} \tag{6.17}$$

where x_t, y_t are in terms of deviations from their means which are given by

$$\bar{x} = \frac{1}{n - \beta(n-2)} \left(\sum_1^n X_t - \beta \sum_2^{n-1} X_t \right). \tag{6.18}$$

These are completely general results for error terms of the simple type considered and do not involve any assumptions about the distribution of the random disturbances ϵ_t. If ϵ_t are normally distributed then we have a maximum likelihood solution.

Comparing the estimate (6.17) with that obtained by our modified transformation procedure of section II we have from (6.11) and (6.12)

$$Y_t - \beta Y_{t-1} = \alpha'_0 + \alpha_1(x_t - \beta X_{t-1}) + \epsilon_t \tag{6.19}$$

where the least squares estimate of α_1 is

$$\hat{\hat{\alpha}}_1 = \frac{\sum_2^n x_t y_t - \beta \sum_2^n x_t y_{t-1} - \beta \sum_2^n x_{t-1} y_t + \beta^2 \sum_2^n x_{t-1} y_{t-1}}{\sum_2^n x_t^2 - 2\beta \sum_2^n x_t x_{t-1} + \beta^2 \sum_2^n x_{t-1}^2} \tag{6.20}$$

where the means are calculated by

$$\frac{1}{n-1} \sum_2^n X_t \quad \text{and} \quad \frac{1}{n-1} \sum_2^n X_{t-1}. \tag{6.21}$$

If we represent the numerator and denominator of (6.20) by A and B respectively we obtain

$$\hat{\hat{\alpha}}_1 = \frac{A}{B} \tag{6.22}$$

so that the estimator given by (6.17) is

$$\hat{\alpha}_1 = \frac{A + x_1 y_1 (1 - \beta^2)}{B + x_1^2 (1 - \beta^2)} \tag{6.23}$$

The reason for this difference is that $\hat{\hat{\alpha}}_1$ ignores the possibility of making use of the first error term u_1, and estimates the regression coefficients using only $(n-1)$ transformed terms. The sum of squares of the $(n-1)$ terms is

$$\sum_2^n \epsilon_t^2 = \sum_2^n (u_t - \beta u_{t-1})^2. \tag{6.24}$$

The first term may be introduced by using the fact that the expected value of ϵ_1^2 given u_1 is

$$E(\epsilon_1^2) = (1 - \beta^2)u_1^2 \tag{6.25}$$

so that

$$s^2 = \sum_1^n \epsilon_t^2 = \sum_2^n (u_t - \beta u_{t-1})^2 + (1 - \beta^2)u_1^2. \tag{6.26}$$

If we substitute for the u's in terms of x and y from (6.11) and minimize in the ordinary way with respect to α_0 and α_1, we again obtain the solutions (6.17) and (6.18). It can be seen therefore that $\hat{\hat{\alpha}}_1$ is an unbiased estimate of α_1 but by ignoring the first term a maximum of one degree of freedom is lost in the transformation procedure as β approaches zero. As β approaches unity the difference between $\hat{\alpha}_1$ and $\hat{\hat{\alpha}}_1$ approaches zero and when $\beta=1$ the solutions (6.17) and (6.20) are identical and the obvious course is to make a first difference transformation.

In the case of multivariate regression the procedure of transforming the variables and applying ordinary least squares analysis provides a much simpler solution than the method indicated by (6.17). The transformation procedure also provides a simpler solution in the case where the autoregressive structure of the error term comprises a linear stochastic difference equation involving two or more lagged terms.

REFERENCES

[1] R. Frisch, *Statistical Confluence Analysis by means of Complete Regression Systems*, Oslo 1934.

[2] T. Haavelmo, "The Probability Approach in Econometrics," *Econometrica*, Vol. 12 Supplement, July 1944, and "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, Vol. 11, 1943, pp. 1–12.

[3] T. Koopmans, "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, Vol. 40, 1945, pp. 448–466, and "Statistical Methods of Measuring Economic Relationships," *Cowles Commission Discussion Papers*, Statistics No. 310. (Mimeographed copy of lectures delivered at the University of Chicago 1947.)

[4] See for example M. A. Girshick and T. Haavelmo, "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations," *Econometrica*, Vol. 15, 1947, pp. 79–110; J. Marshak and W. H. Andrews, "Random Simultaneous Equations and the Theory of Production," *Econometrica*, Vol. 12, pp. 143–205; for a mathematical treatment, T. W. Anderson and H. Rubin, "Estimation of the Parameters of a Single Stochastic Difference Equation in a Complete System," to be published in *Annals of Mathematical Statistics.*

[5] G. Tintner, "The Variate Difference Method," Cowles Commission Monograph No. 5, 1940.

[6] G. U. Yule, "Why do we sometimes get nonsense correlations between time-series etc.?," *Journal of the Royal Statistical Society*, Vol. 89, 1926, pp. 1–64; M. S. Bartlett, "Some Aspects of the Time-Correlation problem in regard to Tests of Significance," *Journal of the Royal Statistical Society*, Vol. 98, 1935, pp. 536–543, and "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series," *Journal of the Royal Statistical Society*, Vol. 8, 1946, pp. 27–41; Galvenius and H. Wold, "Statistical Tests of H. Alfven's Theory of Sunspots," *Arkiv for Matematik*, Astronomi Ochfysik Band 34A No. 24, pp. 1–9; G. H. Orcutt and S. F. James, "Testing the Significance of Correlation between Time Series," *Biometrika*, Vol. 35, 1948, pp. 1–17.

[7] A. C. Aitken, "On Least Squares and Linear Combinations of Observations," *Proceedings of Royal Society Edinburgh*, Vol. 55, 1934/5, pp. 42–48.

[8] D. G. Champernowne, "Sampling Theory applied to Autoregressive Sequences," to be published in *Journal of the Royal Statistical Society.* Series B Vol. 10, 1948.

[9] R. Frisch, *op. cit.*

[10] T. Koopmans, "Linear Regression Analysis of Economic Time Series," Netherlands Economic Institute Haarlem 1937.

[11] G. Tintner, "Some Applications of Multivariate Analysis to Economic Data," *Journal of the American Statistical Association*, Vol. 41, 1946, pp. 472–500.

[12] O. Reiersøl, "Confluence Analysis by Means of Instrumental Sets of Variables," *Arkiv for Matematik, Astronomi Och Fysik*, Band 32A, No. 4, 1945.

[13] R. C. Geary, "Determination of Unbiased Linear Relations between the Systematic Parts of Variables with Errors of Observation," *Econometrica*, vol. 17, 1949.

[14] G. H. Orcutt and D. Cochrane, "A Sampling Study of the Merits of Certain Transformations in Regression Analysis," to be published.

[15] For a general proof, see F. N. David and J. Neyman, "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs*, Vol. II, London, 1938, pp. 105–116; also C. R. Rao, "Generalisation of Markoff's Theorem and Tests of Linear Hypothesis," *Sankhya*, Vol. 7, 1945, pp. 9–16.

[16] H. Cramer, *Mathematical Methods of Statistics*, Princeton 1946, pp. 548–555, and M. S. Bartlett, "On the Theory of Statistical Regression," *Proceedings of Royal Society Edinburgh*, Vol. 53, 1933, pp. 260–283.

[17] A. C. Aitken, *op. cit.*

[18] G. H. Orcutt, "A Study of the Autoregressive Nature of the Time Series used for Tinbergen's Model of the Economic System of the United States

1919–1932," *Journal of the Royal Statistical Society*, Vol. 10, Series B, 1948, pp. 1–53.
[19] J. Tinbergen, "Statistical Testing of Business-Cycle Theories Vol. II; Business Cycles in the United States of America 1919–1932," League of Nations, Geneva, 1939.
[20] Herman Wold, "A Study in the Analysis of Stationary Time Series," Uppsala 1938, pp. 171–174.
[21] *Econometrica*, Vol. 15, 1947, pp. 111–151.
[22] Mimeographed paper distributed by the author and the Cowles Commission for Research in Economics, Chicago.
[23] *Op. cit.*
[24] Richard Stone, "The Analysis of Market Demand," *Journal of the Royal Statistical Society*, Vol. 108, 1945, pp. 286-391.
[25] The relationship between the ratio of the mean square successive difference to the variance and the serial or autocorrelation coefficient for an infinite series is given by

$$r_1 = 1 - \tfrac{1}{2}\delta^2/s^2$$

where r_1 is the first autocorrelation. It can be seen that as r_1 moves from $+1$ to -1 the ratio δ^2/s^2 moves from 0 to 4.
[26] The actual residuals were not published in the paper by Richard Stone but he has very kindly let us have the calculated residuals for 17 equations which include some revised estimates and a few additional relationships (see Table VIII).
[27] For a more detailed discussion of reduced form methods see Girshick and Haavelmo, *op. cit.*, especially p. 85.
[28] J. von Neumann, "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," *Annals of Mathematical Statistics*, Vol. 12, pp. 367–395; B. S. Hart and J. von Neumann, "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Differences to the Variance," *Annals of Mathematical Statistics*, Vol. 13, pp. 207–214.
[29] M. G. Kendall and B. Babington Smith, *Tracts for Computors No. 24*, Cambridge University Press, 1939.
[30] A. C. Aitken, "On Least Squares and Linear Combinations of Observations," *op. cit.*

AOQL SINGLE SAMPLING PLANS FROM A SINGLE CHART AND TABLE*

DONALD J. GREB
Chief, Quality Control Engineer, Minneapolis-Honeywell Regulator Co., Minneapolis, Minnesota
AND
JULIO N. BERRETTONI
Consultant Economist and Statistician
Minneapolis, Minnesota

This paper presents a single chart and table from which AOQL (Average Outgoing Quality Limit) Single Sampling Plans may be determined with ease. These plans yield a close approximation to minimum inspection both for unknown incoming quality and for known average incoming quality unless the variation in quality from lot to lot is extremely small.

AOQL SINGLE SAMPLING PLANS

CHART I and Table II present a set of *AOQL* Single Sampling Plans. Their manipulation is simple. Given *AOQL* and lot size (N) locate on the chart the c-zone of their point of intersection. For example, if $AOQL=1\%$ and $N=1000$, the point of intersection on the chart falls between the two parallel diagonal lines of zone $c=1$.[1] This value of c is the acceptance number and is the ceiling in number of defectives that permits the acceptance of the lot when a sample is used. The sample size corresponding to the value of *AOQL* and c is found in the Table of Sample Sizes (Table II) and for $AOQL=1\%$ and $c=1$ the sample size is 84. The action that follows is to sample 84 from a lot of 1000 pieces and if one or less defective is found accept the lot without further inspection and if more than one defective is found reject the lot for complete sorting. The results to be expected are (1) there is an absolute guarantee that over a series of lots the average per cent defective will not exceed the selected value of *AOQL*, and (2) unless the variation of incoming quality from lot to lot is very small[2] the *AOQL* will be maintained with an amount of inspection that is of practical significance in approximating the minimum inspection which could be obtained if incoming quality from lot to lot were known.

* Acknowledgment is made to Mr. P. M. Brink of Minneapolis-Honeywell Regulator Co. for indispensable assistance with the calculations and for his excellent work in drawing the graphs.

[1] If the point of intersection falls on a line, use the zone directly below.

[2] The word "small" is used here in the sense of being somewhat less than the $\pm 3\sigma$ limits of normal variation.

The limitation of Chart I is that it assumes sample size is small relative to the lot size. If this is not the case, the sample size is larger than need be. However, the limitation is not of a serious nature as the difference between sample size with or without the assumption that n is small relative to N is not of a large order unless the lot size is very small.

These plans are designed for use whenever there is a desire to maintain an average quality over a series of lots. Thus, they may be used advantageously for inspections between operations, departments, subassemblies, receiving inspection, finished products, etc.

THE BACKGROUND OF CHART I[3]

The Derivation of Combinations of Sample Size and Acceptance Number Yielding a Selected Value of AOQL.

The formula for average outgoing quality in terms of the hypergeometric is[4]

$$AOQ = \sum_{m=0}^{m=c}\left(p - \frac{m}{N}\right)\frac{\binom{Np}{m}\binom{N-Np}{n-m}}{\binom{N}{n}} \tag{1}$$

where m is the number of defectives in a sample of size n, c is the acceptance number and p is the per cent defective of a lot of size N. Assume that p is less than or equal to ten per cent and that sample size

[3] The following will serve as a useful list of references: Dodge, H. F. and Romig, H. G., *Sampling Inspection Tables*, John Wiley & Sons, Inc., New York, 1945. Freeman, H. A., Friedman, M., Mosteller, F., and Wallis, W. A., *Sampling Inspection*, McGraw-Hill Book Co., Inc., New York, 1948. Grant, E. L., *Statistical Quality Control*, McGraw-Hill Book Co., Inc., New York, 1946. Hoel, P. G., *Introduction to Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1946. Peach, Paul, *An Introduction to Industrial Statistics and Quality Control*, Edwards and Broughton Co., Raleigh, N. C., 1945. Wilks, S. S., *Mathematical Statistics*, Princeton University Press, Princeton, N. J., 1947. Working Holbrook, *A Guide to the Utilization of the Binomial and Poisson Distributions in Industrial Quality Control*, Stanford University Press, Stanford University, California, 1943.

Churchman, C. W. and Epstein, B., *Tests of Increased Severity*, Journal of the American Statistical Association, Vol. 41, No. 236, December, 1946, pp. 567–590. Dodge, H. F., *A Sampling Inspection Plan for Continuous Production*, The Annals of Mathematical Statistics, Vol. XIV, No. 3, September, 1943, pp. 264–279. Wald, A. and Wolfowitz, *Sampling Inspection Plans for Continuous Production Which Insure a Prescribed Limit on the Outgoing Quality*, The Annals of Mathematical Statistics, Vol. XVI, No. 1, March 1945, pp. 30–49.

Army Service Forces, Office of the Quartermaster General, *Sampling for Quality Control (Supervisor's Edition)*, December, 1945. Navy Department, *General Specification for Inspection of Material*, Appendix X, Standard Sampling Inspection Tables for Inspection by Attributes, April 1946, United States Government Printing Office, Washington, D. C.

[4] Wilks, S. S., *op. cit.*, p. 223 and Hoel, P. G., *op. cit.*, p. 224.

is greater than ten. Then a close approximation is established by the substitution of the Poisson for the hypergeometric distribution. Further assume that N is large relative to n and therefore also to m so that n/N and m/N are considered negligible. Given these assumptions, equation (1) reduces to[5]

$$AOQ = p\sum_{m=0}^{c} \frac{e^{-np}(np)^m}{m!}. \tag{2}$$

Upon maximizing this equation, we have that

$$AOQL = \hat{p}\sum_{m=0}^{c} \frac{e^{-n\hat{p}}(n\hat{p})^m}{m!} \tag{3}$$

and

$$(n)(AOQL) = n\hat{p}\sum_{m=0}^{c} \frac{e^{-n\hat{p}}(n\hat{p})^m}{m!} \tag{4}$$

where $\hat{p}$ is the abscissa value of maximization. Let $a=(n)(AOQL)$, which values[6] for integral variations of c from zero to twelve are presented in Table I. Sample sizes which in combination with c yield selected values of $AOQL$ are readily determined by dividing the $(n)(AOQL)$ values by the given values of $AOQL$. Table II presents these sample sizes.

---

[5] Equation (2) differs from Dodge and Romig's equation of AOQ, which is

$$AOQ = \sum_{M=0}^{N}\left[\frac{N!}{(N-M)!M!}p^M(1-p)^{N-M}\frac{(M-m)}{N}\sum_{m=0}^{c}\frac{C_m^M C_{n-m}^{N-M}}{C_n^N}\right]$$

In words, this equation states that AOQ values, as calculated by the hypergeometric formula of the acceptance from a sample size n of c defectives pertaining to a lot of size N and M defectives, are weighted by the expected binomial frequencies of M defectives, lot size N and average incoming quality equal to p. Thus their assumption is that a lot is a sample from a stream of statistically controlled product varying according to the binomial distribution. The equation can easily be reduced to the summary form of

$$AOQ = p(1-n/N)\sum_{m=0}^{c}\frac{n!}{(n-m)!m!}p^m(1-p)^{n-m}$$

and by substituting the Poisson for the Binomial the equation given for AOQ in their book is obtained (*op. cit.*, p. 48, equation 15). It is to be noted that given our assumption that n/N is small our definition does not differ from theirs and also that with this assumption AOQ is made independent of the binomial form of distribution and of N.

The writers wish to thank Mr. Dodge and Mr. Romig for their kindness in conveying to us by way of correspondence the underlying aspects of their definition of AOQ.

[6] Poisson summation tables of Grant or Molina may be used to calculate $n(AOQL)$ values. Grant, E. L., *op. cit.*, Table G, pp. 542–546. Molina, E. C., *Poisson's Exponential Binomial Limit*, D. Van Nostrand, New York, 1947, Table II.

TABLE I

VALUES OF $n(AOQL) = a$

c	np*	Pa	$n(AOQL)$**
0	1.000	.367879	.3679
1	1.618	.519136	.8400
2	2.270	.604010	1.3711
3	2.945	.659552	1.9424
4	3.640	.698775	2.5435
5	4.349	.728499	3.1682
6	5.071	.751730	3.8120
7	5.804	.770495	4.4720
8	6.546	.786079	5.1457
9	7.297	.799148	5.8314
10	8.055	.810388	6.5277
12	9.590	.828740	7.9476

* Accurate to .0005.
** Accurate to .00005.

TABLE II

TABLE OF SAMPLE SIZES

AOQL—Per cent

c	.10	.25	.50	.75	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	6	7	8	9	10
0	367	147	73	49	36	24	18	14	12	10	9	8	7	6	5	4	4	3
1	840	336	168	112	84	56	42	33	27	24	21	18	16	13	11	10	9	8
2	1371	548	274	182	137	91	68	54	45	39	34	30	27	22	19	17	15	13
3	1942	776	388	258	194	129	97	77	64	55	48	43	38	32	27	24	21	19
4	2543	1017	508	339	254	169	127	101	84	72	63	56	50	42	36	31	28	25
5.		1267	633	422	316	211	158	126	105	90	79	70	63	52	45	39	35	31
6			762	508	381	254	190	152	127	108	95	84	76	63	54	47	42	38
7				596	447	298	223	178	149	127	111	99	89	74	63	55	49	44
8					514	343	257	205	171	147	128	114	102	85	73	64	57	51
9						388	291	233	194	166	145	129	116	97	83	72	64	58
10								261	217	186	163	145	130	108	93	81	72	65
12											198	176	158	132	113	99	88	79

c-p and c-N Zones of Minimum Average Inspection and the Construction of a Minimum Inspection Single Sampling Chart

The formula for average nnmber of pieces inspected per lot (I) is[7]

$$I = n_c + (N - n_c)\left(1 - \sum_{m=0}^{c} \frac{e^{-n_c p}(n_c p)^m}{m!}\right). \qquad (5)$$

[7] The subscripts of n are c and $AOQL$; the omission of $AOQL$ is a matter of convenience in notation.

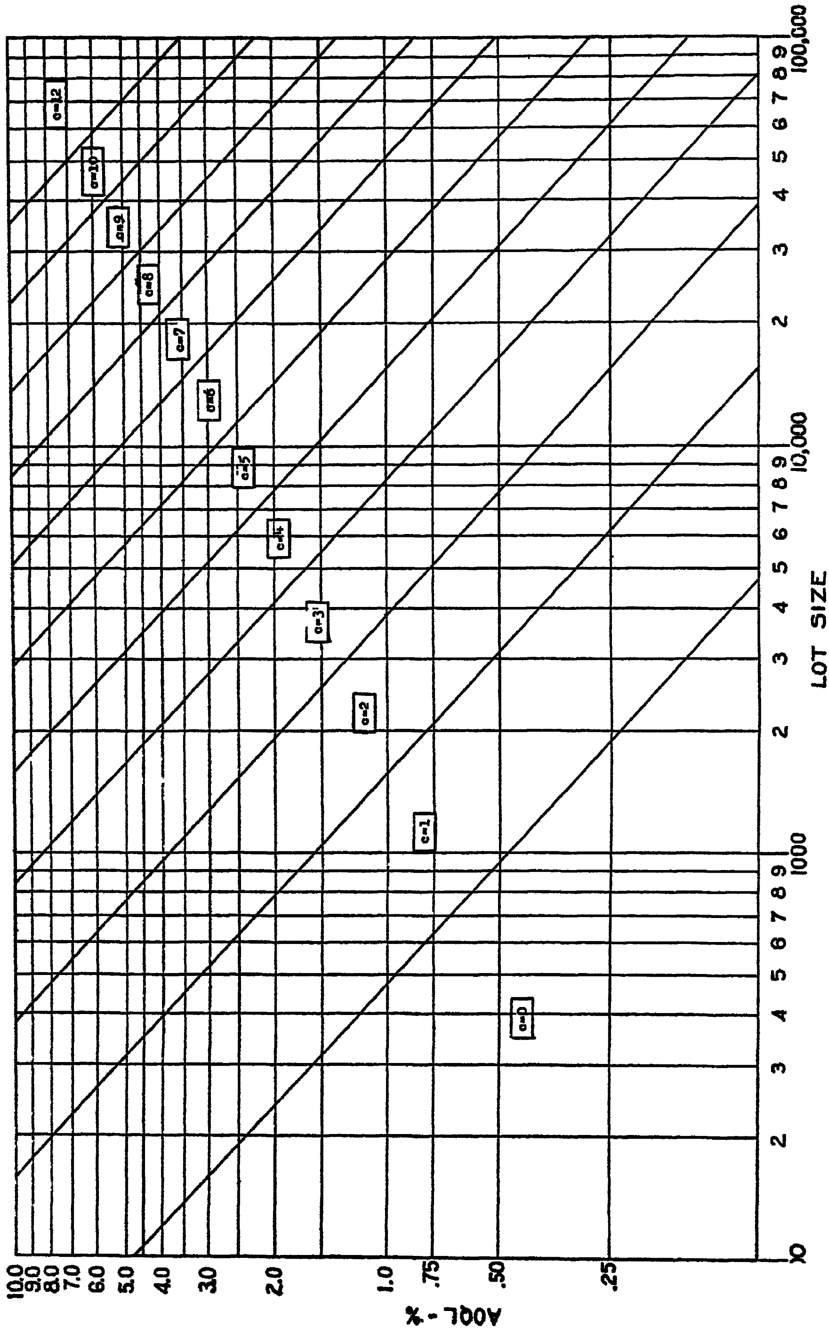


CHART I. *AOQL* SINGLE SAMPLING PLANS.

Assume that N and $AOQL$ are constant.[8] Equation (5) then algebraically characterizes curves illustrated by those in Chart II for the value of c from zero to four, given $AOQL = 2\%$ and $N = 1000$.[9] This chart is significant in three respects. First, it shows that as p increases inspection curves intersect and form zones of minimum average inspection for certain ranges of p. These are the popular c-p zones introduced by H. F. Dodge and H. G. Romig.[10] Second, it is instructive in that as p varies from zero to one, the c factor giving minimum inspection varies parabolically with a maximum of three. That is, as p varies from zero to one, the variation of c forming minimum inspection zones is 0, 1, 2, 3, 2, 1, and 0. This signifies that only sampling plans with these values of c yield minimum inspection. Sampling plans using c equal to or greater than four do not involve minimum inspection with any value of p. Third, it points out that the ratio $p/AOQL = 1$ is contained within the zone formed by the maximum of the c values yielding minimum inspection. Chart III presents only the segments of the inspection curves of Chart II which form zones of minimum average inspection. This curve of Chart III is designated as the c-p minimum inspection curve. If it is assumed that p and $AOQL$ are constant, then equation (5) represents inspection curves forming c-N zones of minimum inspection which are illustrated in Chart IV. In this case, the designation of c-N is attached to the minimum inspection curve.[11]

The equation

$$n_c + (N - n_c)\left(1 - \sum_{m=0}^{c} \frac{e^{-n_c p}(n_c p)^m}{m!}\right)$$
$$= n_{c+1} + (N - n_{c+1})\left(1 - \sum_{m=0}^{c} \frac{e^{-n_{c+1} p}(n_{c+1} p)^m}{m!}\right) \qquad (6)$$

gives values of p and N demarcating the boundaries of c-p and c-N zones of minimum inspection respectively. The following equation, derived from equation (6)

[8] Fixing the value of $AOQL$ determines the values of n_c.

[9] When only c is specified, the corresponding value of n_c is to be understood.

[10] *Op. cit.*

[11] c-$AOQL$ zones also may be derived by holding constant p and N. However, these zones have only theoretical value and are not discussed here. The zones of minimum average inspection may be succinctly analyzed in terms of differences in sample sizes and amount of detailing as expressed by the equation

$$I_{k-c} - I_c = (n_{k-c} - n_c) - \left[(N - n_c)Pr_c' - (N - n_{k-c})Pr_{k-c}\right] \; k = c, c+1, c+2 \cdots \infty, k \neq 2c$$

$$pN = \frac{\dfrac{p}{AOQL}\left[a_{c+1}\sum_{m=0}^{c+1}\dfrac{e^{-a_{c+1}p/AOQL}(a_{c+1}p/AOQL)^m}{m!} - a_c\sum_{m=0}\dfrac{e^{-a_cp/AOQL}(a_cp/AOQL)^n}{m!}\right]}{\sum_{m=0}^{c+1}\dfrac{e^{-a_{c+1}p/AOQL}(a_{c+1}p/AOQL)^m}{m!} - \sum_{m=0}^{c}\dfrac{e^{-a_cp/AOQL}(a_cp/AOQL)^m}{m!}} \tag{7}$$

gives an alternate and easier method of determining the c values yielding minimum inspection. Plotting Np against $p/AOQL$ (Chart V), minimum inspection zones are described for Np and/or $p/AOQL$. Therefore, if N and $AOQL$ are constant, the c number of the c-p zone of minimum inspection is read directly from Chart V. For instance, if $AOQL=2\%$, $N=1000$, and $p=1\%$, then $Np=10$ and $p/AOQL=0.5$ and the coordinate falls in the zone of $c=2$, which result is the same as that of Chart III (or II). If p and $AOQL$ are given, it follows that c-N zones of minimum inspection are obtained, so that if p and $AOQL$ are each equal to 2% and $N=400$, then $Np=8$, $p/AOQL=1$ and the chart reads $c=1$. The same result is found in Chart IV. Therefore, Chart V presents a minimum inspection $AOQL$ single sampling chart for known values of p and N.

Chart V is of further interest in that it summarizes the characteristics of c-p and c-N zones of minimum inspection. In Chart VI the zones through which the dashed curves pass are the $p/AOQL$ zones of minimum inspection for the designated values of N and $AOQL$ and these zones are in direct proportion to c-p zones. It is readily seen, therefore, that c-p zones vary parabolically and that $p/AOQL=1$ is always contained in the zone of the maximum c of minimum inspection. Furthermore, it is noted that the number of c zones and the maximum c vary directly with N. The zones through which the vertical lines of $p/AOQL$ pass are in direct proportion to c-N zones. If $p/AOQL$ is equal to or less than one all zones in Chart V (except that forming the boundary between $c=1$ and $c=0$) converge to a point at $Np=$infinity and $p/AOQL=0$. Thus, in this region as N approaches infinity, c-N zones also approach infinity. If $p/AOQL$ is greater than one, all zones become vertically asymptotic for definite ranges of $p/AOQL$ values so that as N increases c-N reaches a definite maximum value. For example, if $p=1.8\%$, $AOQL=1\%$ then every value of N higher than 724 will have $c=1$ for

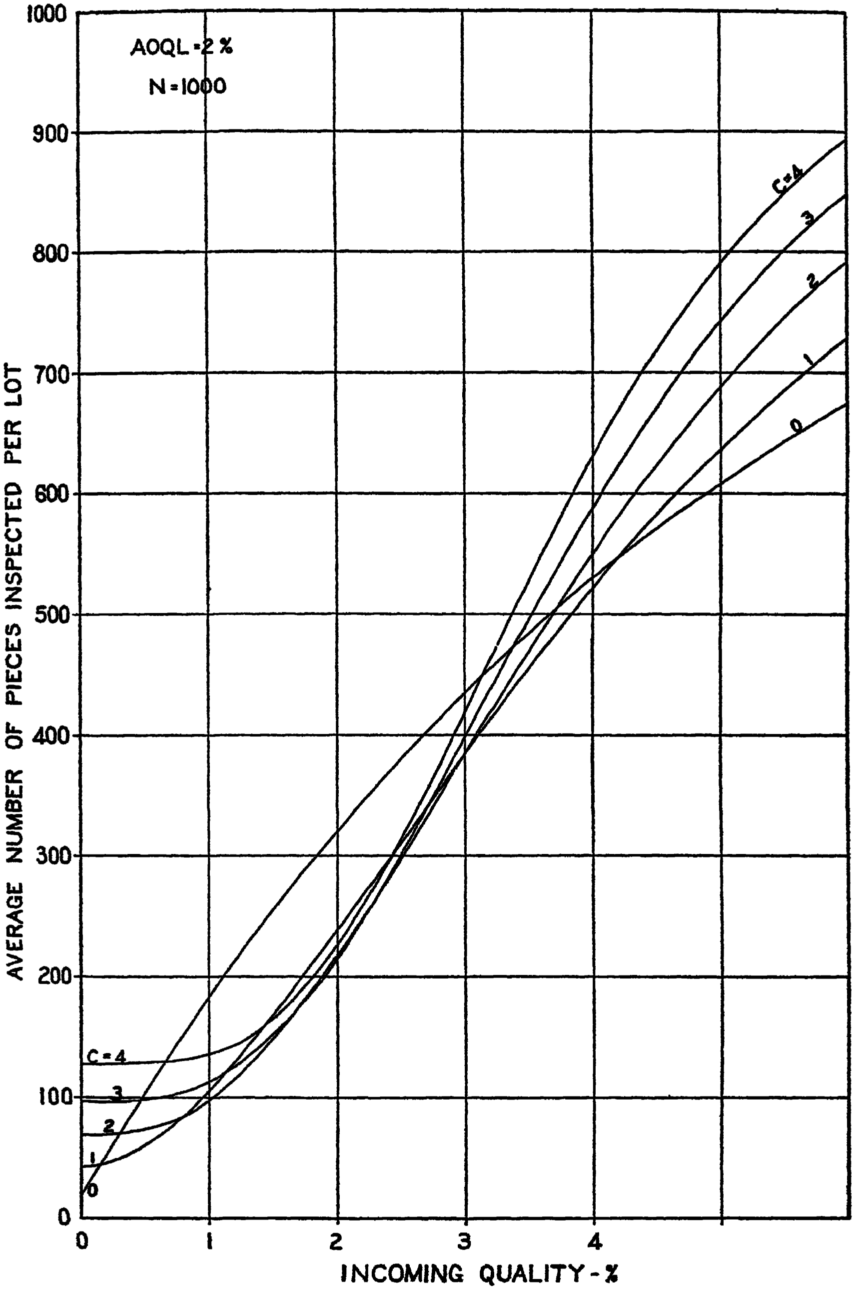


CHART II. EXAMPLE OF c-p ZONES.

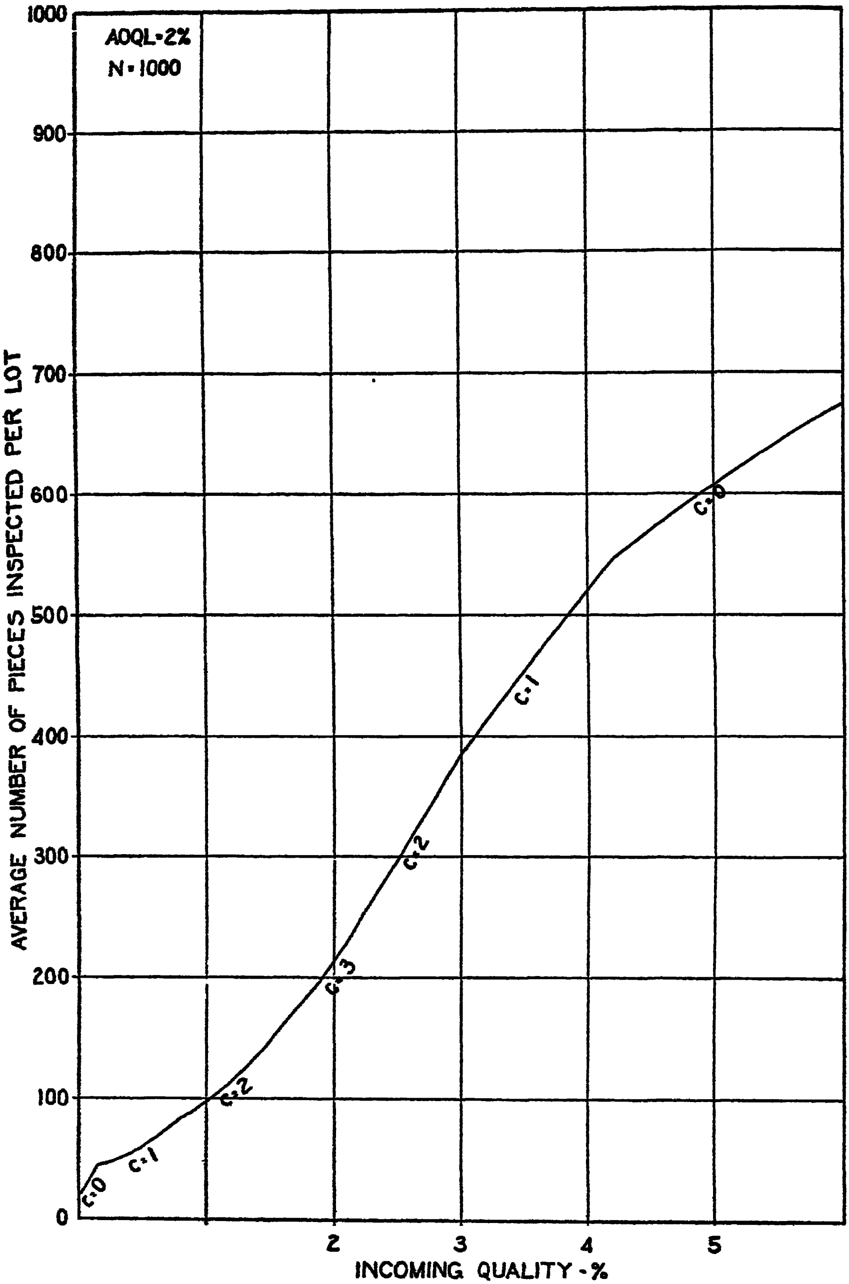


CHART III. c-p MINIMUM INSPECTION CURVE.

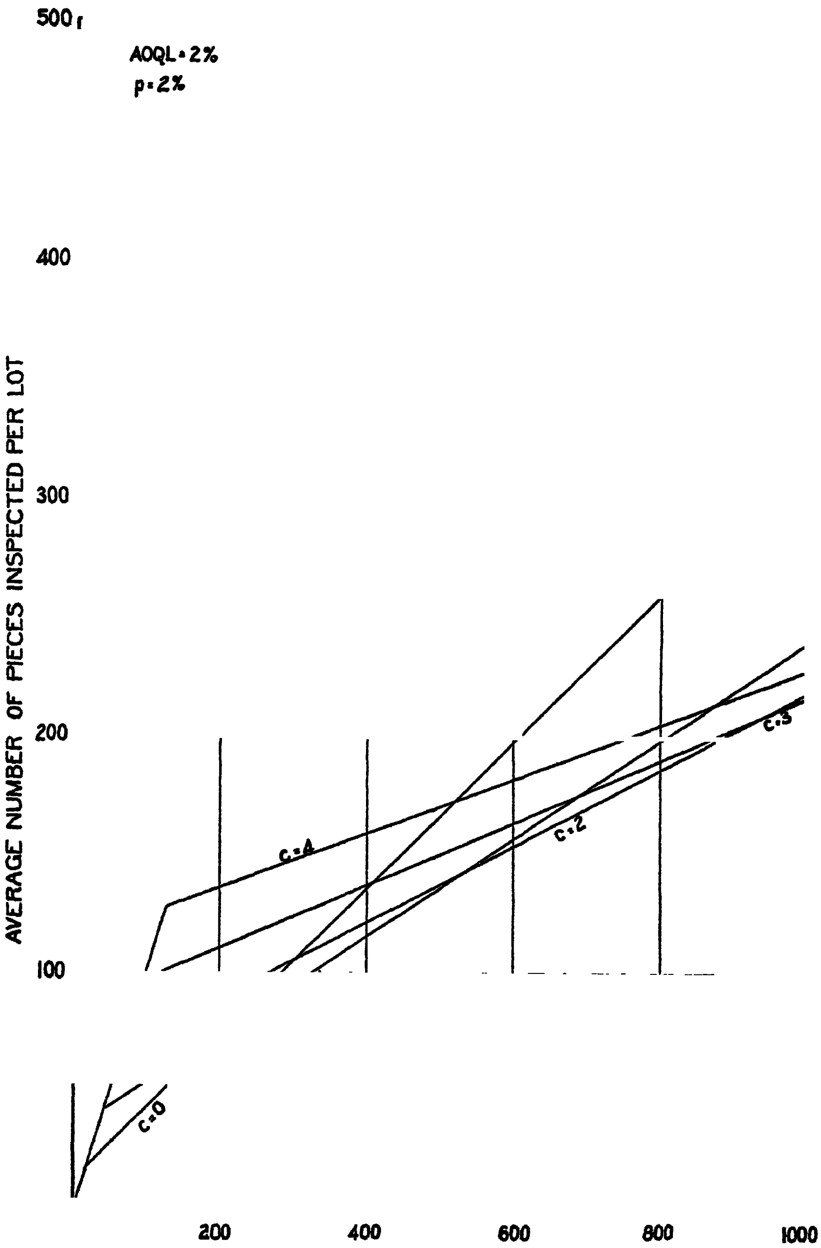


CHART IV. EXAMPLE OF c-N ZONES.

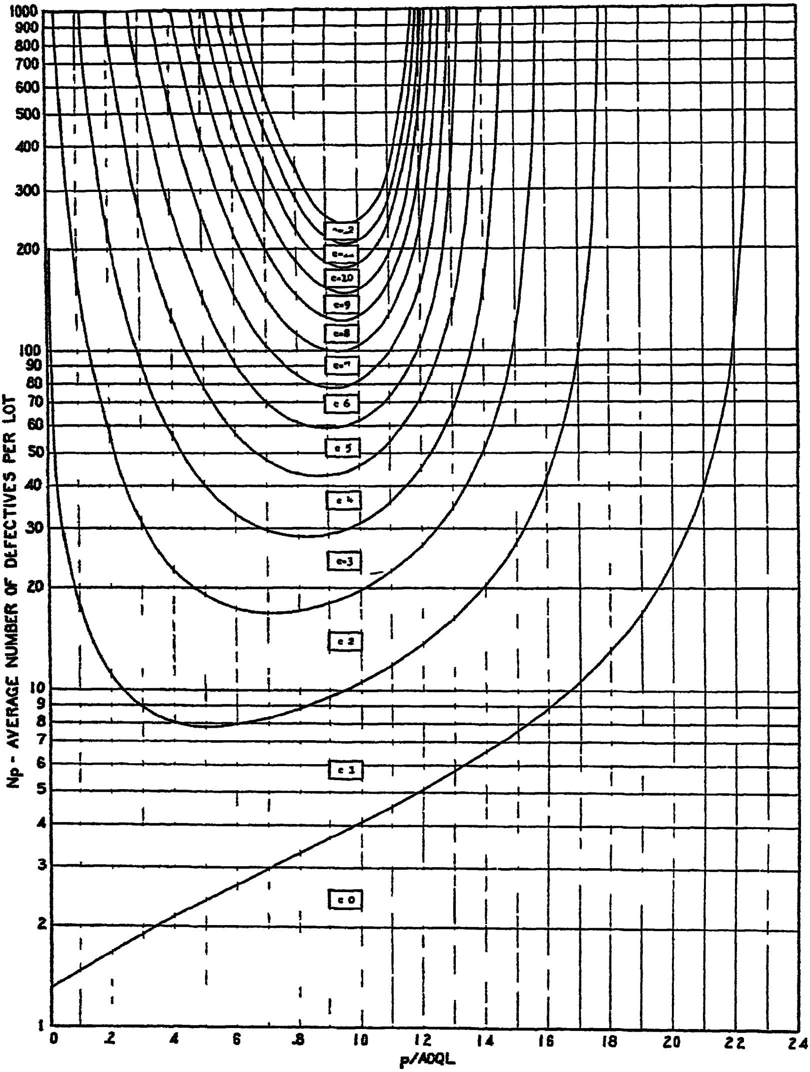


CHART V. *AOQL* MINIMUM INSPECTION SINGLE SAMPLING CHART

minimum inspection since the zone boundaries are asymptotic at values of $p/AOQL$ equal to approximately 1.79 and 2.24+.

AOQL Single Sampling Plans

The selection of c yielding minimum inspection depends on a knowledge of p and therefore Np and $p/AOQL$. In practice these values are always unknown. However, it is known from Chart V that any arbitrary selection of a value of $p/AOQL$ will give a value of c that forms a zone of minimum inspection. It is also known that the maximum c number of minimum inspection can be determined from N if the value of $p/AOQL$ is assumed to be 1. This value of c will lead to minimum inspection for a certain range of p and deviate from minimum for other values of p. Thus, if this value of c is used as a substitute for those based on known values of p, inspection over the entire range of p would approximate minimum. This is shown in Chart VII, which compares the $c=3$ inspection curve ($c=3$ obtained by assuming $p/AOQL=1$ and $N=1{,}000$) and the c-p minimum curve of $AOQL=2\%$ and $N=1000$.[12] Similarly, it is known that the use of the ratio $p/AOQL$ equal to zero gives the smallest c of minimum inspection, namely zero, and that minimum inspection is always obtained for values of p beyond $2.24 \cdot AOQL$.[13] If this value of c is used as a substitute for those based on known p, a second approximation to minimum inspection is obtained for the overall variability of p. Chart VII gives a visual presentation of the approximation when $c=0$ is used. However, because of the rapid convergence of inspection curves to the value of N as p varies beyond $2.24 \cdot AOQL$, or stated differently, because of the large amount of detailing for any value of c beyond $2.24 \cdot AOQL$—never less than 56%—it is of little significance from the economic point of view whether the minimum $c=0$ or some other value is used. The importance of approximation lies in the region of p less than $2.24 \cdot AOQL$. In studying inspection curves of approximation within this region, it has been found that the ratio $p/AOQL=0.5$ gives a better approximation to minimum inspection than any other value of $p/AOQL$.[14] For the entire range of p, with practical significance, the ratio of $p/AOQL=0.5$ leads to the selection of an inspection curve which best approximates the c-p minimum inspection curve.

[12] After the ratio $p/AOQL$ has been assigned, the given value of $AOQL$ (e.g. 2 per cent) is used only to determine the sample sizes.

[13] See Chart V.

[14] To obtain the same inspection curves of approximation given by $p/AOQL=0.5$, $p/AOQL$ must necessarily vary in the region of $p/AOQL>1$ because of the asymptotic nature of the c zones in this region.

The selection of c-values based on the assumption of $p/AOQL=0.5$ is made easier by constructing Chart I. This chart eliminates the requirement of Chart II of calculating Np. Since from equation (7)

$$N=\frac{\dfrac{1}{AOQL}\left[a_{c+1}\sum_{m=0}^{c+1}\dfrac{e^{-a_{c+1}p/AOQL}(a_{c+1}p/AOQL)^m}{m!}-a_c\sum_{m=0}\dfrac{e^{-a_cp/AOQL}(a_cp/AOQL)^m}{m!}\right]}{\sum_{m=0}^{c+1}\dfrac{e^{-a_{c+1}p/AOQL}(a_{c+1}p/AOQL)^m}{m!}-\sum_{m=0}^{c}\dfrac{e^{-a_cp/AOQL}(a_cp/AOQL)^m}{m!}} \tag{8}$$

and since N, $p=AOQL/2$, c, $(n_c)(AOQL)=a_c$, and $(n_{c+1})(AOQL)=a_{c+1}$ are given, we have that

$$N=\frac{A}{AOQL} \tag{9}$$

$$\log N=\log A-\log AOQL \tag{10}$$

where

$$A=\frac{a_{c+1}\sum_{m=0}^{c+1}\dfrac{e^{-\frac{1}{2}a_{c+1}}(\frac{1}{2}a_{c+1})^m}{m!}-a_c\sum_{m=0}^{c}\dfrac{e^{-\frac{1}{2}a_c}(\frac{1}{2}a_c)^m}{m!}}{\sum_{m=0}^{c+1}\dfrac{e^{-\frac{1}{2}a_{c+1}}(\frac{1}{2}a_{c+1})^m}{m!}-\sum_{m=0}^{c}\dfrac{e^{-\frac{1}{2}a_c}(\frac{1}{2}a_c)^m}{m!}} \tag{11}$$

and is a constant. Thus, in logarithms the boundary lines between lot sizes and $AOQL$ are parallel, straight and negatively sloped.

Comparison of Charts I and V Given Knowledge of Average Incoming Quality ($\bar{p}$)

Chart V yields minimum inspection if the value of p for an incoming lot is known. Therefore, Chart V is useful in practice if $\bar{p}$ is known and if the variation of p is almost completely within a single zone of minimum inspection. In general, this will require a variation considerably less than the normal plus or minus three sigma limits of variability. Consequently, unless the variation of p is very small, Chart I gives a better approximation to minimum inspection because of the parabolic behavior of the c-p zones.

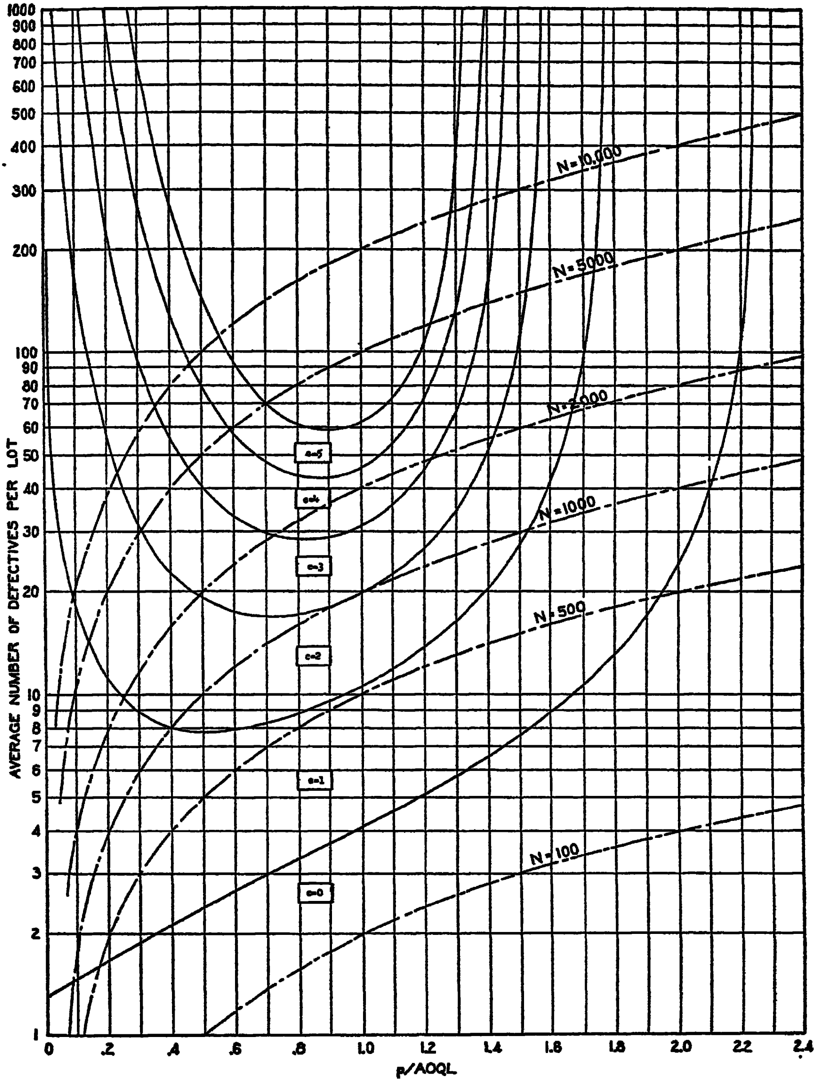


CHART VI. CHARACTERISTICS OF $c-p$ AND c-N MINIMUM INSPECTION ZONES.

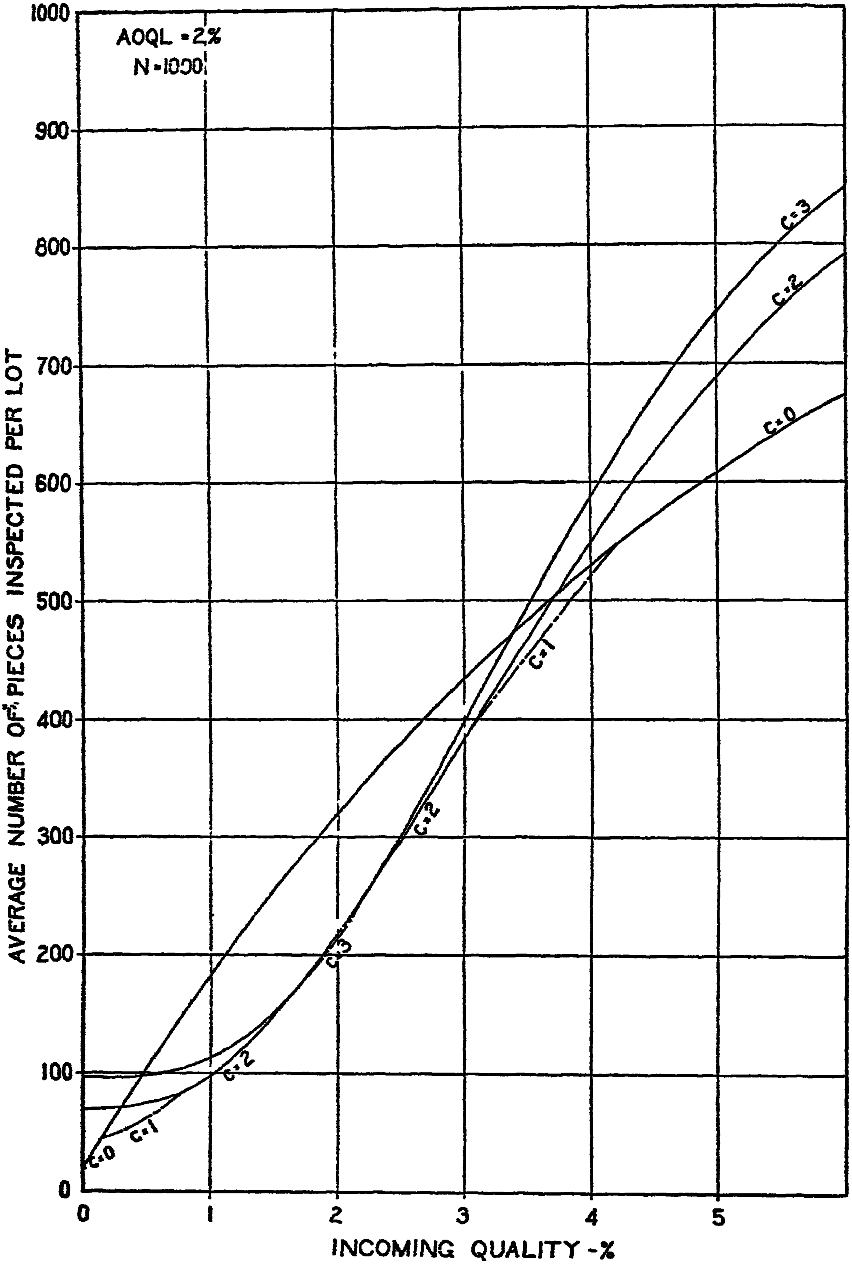


CHART VII. COMPARISON OF INSPECTION CURVES OF $c=3$, $c=2$, AND $c=0$ TO c-p MINIMUM INSPECTION CURVE.

ON MEASURING LANGUAGES

STUART C. DODD, PH.D.
University of Washington, Seattle

This paper proposes ten criteria by which the suitability of any language for use as an international language might be measured. These criteria fall into two classes. The first three are criteria of familiarity—that is, they measure the extent to which a candidate language is already familiar to the people who would have to learn it. The remaining seven are criteria of excellence, and are intended to rate languages according to such properties as their freedom from local idioms, from exceptions to the rules of grammar, from inflections, and so on. Such criteria have three purposes. First, they would rank the candidate languages by familiarity and excellence. Second, they would diagnose weaknesses in each candidate: from this diagnosis a living language could be simplified towards the ideal regularity of an artificial language, while preserving more of familiarity to the world's population than an artificial language possesses. Finally, they would indicate any progress that the world may be making from decade to decade towards achieving a single language.

THE PROBLEM OF an international auxiliary language has become in part a problem of selecting it from among the three hundred candidates which have been proposed in the last seventy years. To select the best candidate requires prior agreement on what is "best." What are the criteria which specify the "best"? This paper proposes ten criteria. It further proposes ten indices which measure the degree to which each criterion is satisfied by a given candidate language. A weighted sum of these indices can then rank the candidates into a relative order.

The criteria for the best world language may be put into two classes —the practical and the ideal. These are also called the natural vs the schematic types when applied to artificial languages. They specify, respectively, what is most likely to be adopted by the world and what is intrinsically the most excellent as a language. These will be referred to here as the criteria of familiarity and the criteria of excellence.

For it should be obvious that the most practical proposal is one which involves the least change or the least amount of new learning for the world's population. Thus the candidate language which has the largest proportion of elements which are already familiar to the maximum number of people will encounter least resistance. We can measure the degree of familiarity of a candidate language and thus make a crucial comparison to prove which candidates are most practical.

THE FAMILIARITY CRITERION

What language is most familiar to the users of any one language in that it has the largest percentage of its words, grammatical forms or other elements the same between their language and the candidate world language? In order to measure this, several indices, varying in completeness are as follows: A first index of familiarity, which may be labelled F_1, is calculated by taking as a first step a representative sample of the elements of the candidate language. One such sample might be the 1000 semantic words which occur most frequently, as determined from a semantic word count. (A "semantic" word is defined as a word or phrase with a unit meaning i.e. "to look out for" meaning "to protect.")

The next step is to assign to each of these thousand "most frequent" words, which serve as a representative sample of the candidate language, a value of 1 if it is exactly the same in the national language and a value of $\frac{1}{2}$ if it is partly the same (as in having a root or an affix in common). These unit or half unit values are added up and, since they will give a thousand points at maximum, this total will serve as a percentage of common vocabulary between one candidate language and one national language. This percentage is F_1.[1]

Next there will be other values of F_1, one for each national language paired with each candidate language. That is, for one candidate language there will be as many F_1's as there are national languages or important groups of national languages deserving consideration in the world. This number of F_1 indices will then be multiplied by the number of candidate languages i.e. the number of languages for which research provides these data and which are considered important enough to be likely candidates for a world language. It is obvious that this is an immense project of research for many scholars for many years.

These F_1 indices of familiarity next must be combined into a net index of familiarity for each candidate language for the whole world. That is, the F_1 index for one candidate language must be weighted or each multiplied by the number of people speaking the national language corresponding to that index. This gives greater importance to the familiarity index of a language spoken by a 100 million people than to one spoken by one million people. From this democratic process of weighting each index by the population to which it applies, there will result a single net index of familiarity, which we may call F_2, for each candidate language. These indices will rank the different candidates in

---

[1] $F_1 = (\Sigma V/N)$ where V = value of 1 or $\frac{1}{2}$ and N = number of words in sample studied.

order of familiarity to the world. The indices reveal the languages at the top of the list which deserve further study and also reveal those at the bottom of the list which may be dropped from further consideration.

A number of problems of method will have to be solved in computing these indices. For example, in determining the number of people speaking each national language, in order to fix upon a weighting of its F_1 index, allowance must be made for bilingual people or that fraction of a population which may speak more than one language. Such persons might be counted as ½ for each of the two languages they may speak thus giving them a total weighting the same as for any person speaking but one language. Again a further refinement in the indices might be to weight each word in calculating the F_1 index in proportion to the frequency of occurrence of that word. A priori, it seems probable, however, that if the thousand most frequently used words are taken as the sample in calculating F_1, differences in the frequency of individual words would not greatly change the relative size of the F_1 indices.

A third problem is whether to take the total population speaking a given national language as a weighting factor in calculating F_2 or whether to take some part of it which is more relevant for international purposes. Thus the *literate* part of each national population is probably a more suitable number to take as weighting coefficient. This index might be called F_3, measuring the degree of familiarity to literate people. For the literate population represents those who are communicating in international affairs more adequately than the host of illiterates. To include the illiterates would give the 400 million or more illiterates of China or India an importance greater than all the Western European nations combined. To weight each nation in proportion to its literate population would probably be fairer basis, since part of learning an international language is learning its written forms. An index of familiarity should apply in part to the people who have already learned some written form of language and might have to unlearn and *relearn* an international language more than to the people who have learned no written form and to whom learning a new word would be little more difficult from learning their own national written forms.

THE EXCELLENCE CRITERIA

In analyzing next, the excellence of any language for international communication the following criteria are proposed as hypotheses. Some combination of criteria such as these would define what is meant by "the most excellent language" for international communication.

The eight proposed criteria of excellence are that: its sentences should be idiomless and ordered in wording; its words should be univocal in meaning, flectionless, phonetic in spelling and unique in pronunciation; its letters should be unique in sound and shape.

A world language should be idiomless. It should not have phrases which are local and peculiar to one nation and cannot be literally translated into other languages. A world language should have all its phrases so logical as to enable literal translation into any national language. To measure the freedom from idioms of any language, a list of its idioms as found in a frequency count of a representative sample of perhaps a million words of prose should be made. The index E_1[2] would be the ratio of the million words of prose examined divided by that million plus the number of idioms (including repetitions) found in that representative body of prose. If many idioms are found, this ratio would be a small percentage. It would become a 100%, indicating a language entirely free from idioms, only when no idioms are found.

To detect an idiom, three tests are available. The first test is the definition of an idiom as a phrase different in meaning from its constituent words. Another test is to try translating each phrase into each of some dozen other representative languages and see whether that phrase can be translated literally. Another test is to see if each phrase can be expressed in the symbols of modern Symbolic Logic. This new science, grown up in the last half century, develops an algebra for words and sentences, so that these qualitative symbols can be handled in equations with all the precision of mathematics.

A world language should have the order of the words in its sentences obey rules without exception. The rule of course may be very rigid; or very flexible as in stating that certain words may occur anywhere in the sentence depending on the emphasis desired. Ideally, it is possible to conceive of a language in which all word order is determined by one rule such as that "modifiers follow that which they modify." This rule would mean that a verb followed the subject and that the object of the verb followed the verb whose meaning it completes. This rule would mean that adjectives followed the noun they modify and adverbs follow the verbs they modify and every phrase or clause follows whatever it modifies.

The index, which we may label E_2, which measures the excellence of a sentence in having the order of its words abide by rule could be com-

[2] $E_1 = (N/(N+M))$ where N = number of words in sample studied; M = number of idioms.

puted as a ratio of the number of words in a representative sample of prose, (perhaps one million words), to the number of these words when each is multiplied by its "frequency rank". This "frequency rank" needs explaining. It is determined as follows: For each word in a sentence the rule that determines its position in the sentence is decided upon. The frequency of occurrence of each rule must then be determined in a large sample of prose, and the rules put into a rank order so that the most used rule will be given a rank of 1, the next most used rule will be given a rank of 2, etc. Each word is then multiplied by the rank (whether 1, 2, 3, etc.) of the rule which governs its position in the sentence. This multiplication by a rank is "weighting" the word according to the frequency rank (in this case the frequency rank of the rule governing the position of the word in the sentence). By this index,[3] if there is but one rule for all words the weighting factor is 1 and the index will be a 100%, as it will be a million words divided by a million words. If, however, a second rule appears then some of the words in the denominator of the index will be multiplied by two and the index will be less than 100%. If a third rule appears the index will become still smaller in proportion to the number of words covered by that third rule. Thus this index becomes smaller in proportion as the number of rules becomes greater.

A high index, therefore, measures simplicity of language in this respect and a low index measures its complexity or irregularity. The index also is proportional to the frequency with which each rule occurs in the representative sample that is studied. It should be obvious that having a definite word order makes sentences which are clear and unambiguous in meaning. If the order of words in a sentence always follows some rule, there is little possibility of different people interpreting the sentence in different ways. Thus a rule-abiding order of words is an objective way of measuring and controlling the degree of ambiguity in the sentences of a language. This is especially so in a language whose words are not inflected (as explained below).

A world language should have words which are uninflected. This criterion means that no word should ever change its form to express a grammatical inflection such as masculine or feminine gender, persons, or number, tense, voice or mood of a verb or degrees of an adjective. This is the trend of evolution of language. Languages grow up with these grammatical inflections in primitive thinking as when

---

[3] $E_2 = (N/\Sigma R_2)$ where N = number of words in sample studied; R_2 = frequency rank of rule governing the position of each word.

man ascribed masculine and feminine gender to all nouns, simply because man thought of his own difference in sex as existing in everything else around him. But as people developed towards greater maturity and flexibility in language they dropped these grammatical inflections. Some of them are entirely unnecessary. Others are expressed in uninflected "particles" such as the prepositions and conjunctions and adverbs like "to," "as," "and," "or," "not," etc. Chinese has gone furthest in developing a completely uninflected language of root words which can be flexibly combined in different orders to make a great variety of meanings.

The flexibility of uninflected words can be compared to the flexibility of the alphabet where the clumsy symbols for whole syllables were replaced by a few letter symbols for elemental sounds. These letters can be flexibly combined to make any word in any world's language. Somewhat similarly, root words and particles yield more flexible sentences with a greater range of possible meanings than inflected words can do.

To measure the degree to which a language has progressed towards the ideal of complete absence of inflections, an index, which may be called E_3, may be calculated from the same representative body of prose of perhaps a million words which may be used for calculating most of these indices discussed in this paper. The formula for the index of freedom from inflections is the ratio of one million words divided by those million words each weighted by its "frequency rank of inflections". This "frequency rank of inflection" is determined in a way similar to the frequency rank of rules in the preceding index, E_2. To get it, the number of times each inflection occurs in the million words is counted, and the frequency of the inflections with one grammatical meaning are given the ranks of 1, 2, 3, etc. Each word is then multiplied by one if it is uninflected, by 2 if its first inflection is the most frequently occurring one, by 3 if its first inflection is the next most frequently occurring one, by a weight of 4 if its first inflection is the next most frequently occurring one, etc. If the word has more than one inflection, it will be multiplied by more than one such rank. By this index,[4] a language will be a 100 per cent flectionless only when it uses root words and particles, only. It will be less than a 100% flectionless in proportion as:

a. It has many words which are inflected

[4] $E_3=(N/\Sigma R_3)$ where N = number of words in sample studied, and R_3 = frequency rank of each inflection of each word.

b. The inflected words are frequent in occurrence, and
c. There is more than one inflection to express one grammatical meaning.

Thus a language which has four conjugations for its verbs instead of one conjugation will have a larger weighting in the denominator of the index and, therefore, a lower index of excellence in respect to being flectionless.

A world language should be phonetic in spelling. This criterion of excellence that every word should be spelled exactly as it is pronounced implies the criterion mentioned below that every letter should represent only one sound. When the words of a language are spelled as they are pronounced, learning to read that language becomes very simple. If there is much literature and reading matter in one's environment, a child will learn to read without schooling as automatically as he learns to speak by merely being surrounded by people using the written language and by his wanting to know what others are writing and to write things himself. A phonetic spelling is perhaps the greatest aid to make the population a 100% literate. All languages which use letters were phonetically spelled at one time of course, but in the case of many languages the spelling of a previous century has become standardized while the pronunciation has changed. Another source of unphonetic spelling, however, is that there are more sounds in a language than letters, so that some letters will be used to mean more than one sound. Thus English uses 40 sounds, but has only 26 letters in its alphabet with a result that its irregularity of spelling is greatly increased.

To measure the degree to which a language is phonetic in spelling an index of this criterion of excellence, which we may call E_1, may be defined by a ratio calculated from a large sample of perhaps 100 thousand letters as they occur in the representative sample of prose referred to above. The index might be a 100 thousand letters divided by the number of those letters when each one is multiplied by its "frequency rank of pronunciation." This frequency rank of pronunciation is again calculated similarly to the frequency rank of rules in E_2 or frequency rank of inflections in E_3. To calculate it the frequency with which each pronunciation of each letter recurs must be counted. Then for any one letter, its most frequent pronunciation is given a rank of 1. Its next most frequent pronunciation is given a rank of 2 and so on. Each letter in the denominator of the ratio is multiplied by its rank and these products are added to make the denominator of E_4. By this index,

a language will be 100% phonetic only when each letter has one pronunciation and when every word is spelled in a single and phonetic way. The phonetic index E_4,[5] of a language will decrease in proportion as its letters, as they occur in words with current spelling, have more than one pronunciation.

A world language should have words which are univocal in meaning. This criterion of excellence means that every word should ideally have only one meaning and every meaning should have only one word to symbolize it. There should be no words with multiple meanings nor should there be any synonyms which mean exactly the same. (Synonyms with slightly different meaning are desirable to express shades of differences in meaning and to make a language rich, but words between which no differences in meaning can be detected are merely confusing.) This is a fundamental principle of symbolism—that each symbol should represent one and only one "referent" or meaning. Obviously, our living languages as they have grown up in folk usage have acquired multiple meanings for many of their words. Only artificial languages such as Esperanto approach the ideal of "one word, one meaning" as they can start out afresh by assigning a word or phrase for every meaning listed in the dictionary.

To measure the excellence of language in respect to its words being unique in meaning, an index, E_5, may be defined as a ratio calculated from the same representative sample of a million words of prose which has been used previously. This index is one million words divided by the number of those words when each is multiplied by its "frequency rank of meaning." This frequency rank of meaning is similar to previous frequency ranks. It would require a semantic word count, i.e. a count of the frequency of occurrence of each meaning (as listed in the dictionary for each word) in the million word sample of prose. (See Eaton's Semantic Frequency List for English, French, German and Spanish.) Each meaning of each word will be given a rank of 1 if it occurred most frequently, of 2 if it occurred next most frequently and so on. Each word would be multiplied by this frequency rank and all these products would be added up to get the denominator of E_5.[6] Since no complete semantic word counts have been made as yet in the world to the author's knowledge, (although a scientific committee is at work on this in the United States) a similar index of uniqueness of meaning of

[5] $E_4 = (N/\Sigma R_4)$ where N = number of letters in sample studied; R_4 = frequency rank of pronunciations of each letter.

[6] $E_5 = (N/\Sigma R_5)$ where N = number of words in sample; R_5 = frequency rank of meaning of each word.

words may be $E_{5A} \cdot E_{5A}$ might be the number of words in the most complete dictionary of a language divided by the number of meanings listed in that dictionary. This variant index is easily computed but has the disadvantages of giving great weight to unusual and archaic meanings with which a language may be burdened and ignores the important factor of the frequency of the use of a word with multiple meanings.

The world language should have uniform pronunciation everywhere. This sixth criterion of excellence means that there should be no difference in different countries in the way any word of the world language is pronounced. By means of the international standardized phonetic alphabet the standard pronunciation of every word can be fixed. Phonograph records and radio recordings can also fix the pronunciation. Some people may comment that since languages have changed their pronunciation in the past would not the new international language also change as a whole or in regional dialects? This is highly improbable as the modern forces such as radio and other agencies of mass communication would increasingly tend to unify and standardize and preserve pronunciation. Dialects grow only where people are separated with little communication between them.

To measure the degree to which any candidate language approaches this ideal of universally uniform pronunciation an index of uniformity of pronunciation, E_6, may be developed. For this index, a survey of perhaps a million words of oral speech would be needed. In this survey, a sample of persons representative of the various regions, social classes, etc. within each national language might be asked to read standardized prose into a recording machine. From these recordings, the frequency of each pronunciation of each word could be counted, and each pronunciation of a word given a rank. Then the index E_6,[7] would be that million words divided by the sum of those words when each has been multiplied by its frequency rank of pronunciation. This index, like the previous ones, becomes 100%, showing complete uniformity of pronunciation, when the rank of every word is one so that the index is one million divided by one million. In proportion as there is more than one pronunciation for each word, the denominator increases and the index of uniform pronunciation shrinks. For example, if there were two pronunciations only on the average for every word the ranks of 1 and 2 would occur equally often as weights in the denominator which would then have the average value of 1.5, giving an index of one million divided by a million and a half which is 67% of uniform pronunciation.

---

[7] $E_6 = (N/\Sigma R_6)$ where N = number of words; R_6 = frequency of pronunciation of each word.

Again, if there were, in general, three pronunciations of every word so that the ranks of 1, 2, and 3 occurred about equally often, then the denominator would be twice the size of the numerator and the degree of the uniformity of pronunciation would be only 50%.

A world language should have every letter unique in shape and sound. These two final criteria of excellence of any language apply only to its written and printed forms. The second means that every letter should have only one pronunciation and every elemental sound in the language should have a letter to represent it. The index to measure this is included in the index of phonetic spelling, E_4 above.

For each letter to be unique in shape it means that every letter would have one and only one visual form, regardless of whether it occurs in print or in hand writing, or whether at the beginning of the word (where capitals are used in some languages), or in the middle or end of a word. Thus English has four forms for many of its letters and Arabic has three forms for many of its letters. To measure the degree of uniqueness of shape of letters more exactly a seventh index of excellence, E_7, may be defined as the ratio from a sample of a hundred thousand letters in the representative samples of written and printed prose. This numerator would be divided by the sum of those letters each multiplied by its frequency rank of shape. This frequency rank of shape, like the preceding frequency ranks, would be determined from a count of the frequency of each shape of each letter. Putting them into rank order and multiplying each letter by its rank and adding these products gets the denominator of the index E_7.[8] It will be 100% only when every letter (including its connection to an adjacent letter) has only one shape.

Seven indices of excellence for any language have been defined above. The next scientific step is to combine them into a single index of excellence for any one candidate language. There are various possible ways of combining them. The simplest way is to draw a profile graph. This means to draw a column showing the percentage value of an index and placing the seven columns for the seven indices of one candidate language side by side. The broken line across the tops of these seven columns is the "profile" for that language. By drawing and superposing profiles for the different candidate languages it might be obvious that one or two are far superior in most respects to all the others (or possibly far inferior to the others and so may be dropped from further consideration).

---

[8] $E_7=(N/R_7)$ where N = number of letters in sample studied and R_7 = frequency rank of shape of each letter.

If the profile, however, shows several candidates with overlapping profiles a more exact method of combining the seven indices into one must be used. The simplest way is to get the simple unweighted average, E_A, by adding them together and dividing their sum by 7. This gives a simple average index of excellence for one language permitting its excellence to be compared with the excellence of other candidate languages. If more refined weighting is desired, it can be secured by having panels of judges who are experts in the science of language distribute a 100 points to the 7 criteria so as to show the relative importance of each. The average number of points assigned by the judges to each criterion would then be a weighting factor for that criterion. This weighting factor for each criterion would be multiplied by the value of its index before adding the 7 indices and dividing to get the weighted average index of net excellence for one language, E_W.[9]

Still more refined weighting schemes could be developed such as one based upon the number of man-hours, or the amount of human energy, required to learn and use whatever each index measures. Thus if unphonetic spelling adds 20% of letters to the words of the language in general then the writing, typing, and type setting of that language requires 20% more time than a language having phonetic spelling. Similarly, the number of hours required on the average to learn the irregular flections of a language compared with the number of hours required to learn an otherwise equivalent but flectionless language would yield a weighting factor for the third criterion dealing with flections.

THE PURPOSE OF THE CRITERIA

As a result of the researches outlined above there would be an index of familiarity and an index of excellence (such as F_3 and E_W) for each of the languages, whether an artificial or a living one, which are candidates to become the auxiliary world language. These indices will serve three purposes. First they would rank the candidate languages and tell which was the most familiar and the most excellent. Thus the problem of choosing the "best" world language would find a scientific answer (based on rules in which the subjective element has been minimized).

Secondly, these indices would diagnose and measure weakness and the degree of strength in each candidate language whether in its freedom from idioms, its regularity in word order, its freedom from flections its phonetic spelling, its uniqueness of meaning of words, its uniformity

[9] $E_W = (\Sigma wE/w)$ where E = each of the preceding indices of excellence E_1 to E_7 in turn; W = weight of each index.

of pronunciation or its uniqueness of the sound and the shape of its letters. From this diagnosis any living language could be simplified towards the ideal regularity of an artificial language while preserving more of familiarity to some part of the world's population than artificial languages possess.[10]

A third purpose of these indices is to measure any progress that the world may be making from decade to decade towards achieving a single world language. The relative degree of gain among the rival languages may be measured from period to period partly by technics of representative sampling as in Gallup polls. The degree of a person's knowledge of a language, averaged for a population, must be also included in any accurate measurement. Any trend, however, slight, towards a single language eventually sweeping the field and becoming the sole auxiliary language would be shown, and its spread could be facilitated.

---

[10] This has been done for English. The resulting "Model English," constructed by the author, has the regularity of the most ideal artificial language coupled with greater familiarity to more people than any rival national or synthetic language. Its indices of excellence are all 100% and its indices of familiarity are well above all rivals. This gives Model English first rank by all criteria for a world language.

CONFIDENCE LIMITS IN THE NON-PARAMETRIC CASE

GOTTFRIED E. NOETHER
Columbia University

The purpose of this article is to give a survey[1] of certain methods available for finding confidence limits when nothing is assumed about the population from which a sample has been drawn except possibly that it has a continuous distribution. The following three cases are treated: a confidence band for the unknown cumulative distribution function, a confidence interval for the proportion of a population for which the variate is smaller (or larger) than a given value, and confidence intervals for quantiles. The results have been known for many years, but have often been accessible only to those who were able to follow rather involved mathematical arguments. It is the purpose of this paper to state certain important results without referring to any mathematical proofs.

Introduction. In the application of statistical theory to practical problems the normal distribution occupies a predominant position. If we can assume that the observations which we have taken have come from a normal population a great many of our troubles are over. However it will often happen that we do not know much more about the parent population than is supplied by the sample itself. What can we do then? The arbitrary assumption of normality may obviously lead to completely wrong conclusions. Having only a scant knowledge about the parent population we are forced to make very broad assumptions. Thus in many cases it may be reasonable to assume that the unknown cumulative distribution function (cdf) is continuous. This will be our assumption in all that follows unless stated otherwise. In contrast to the parametric case when the form of the cdf is supposed to be known except for the values of a finite number of parameters this is often referred to as the non-parametric case, since a finite number of parameters are not sufficient to determine the distribution completely.

One of the important problems in the parametric case is the estimation of the unknown parameter—for simplicity we assume that there is just one—by a confidence interval. The question arises if the idea of a confidence interval can be extended to the non-parametric case. Such an extension was made by Wald and Wolfowitz [1].

---

[1] I should like to thank Professor Wolfowitz for pointing out the need for such a survey

Confidence Bands for Unknown Cumulative Distribution Functions. Before going into the non-parametric case it may be well to review briefly the basic idea underlying confidence intervals in the parametric case. Let the random variable X have the cdf $F(x, \theta)$, i.e., $P\{X \leqq x | \theta\} = F(x, \theta)$, where $P\{A | B\}$ as usual denotes the probability of A computed under the assumption that B is true. The functional form of $F(x, \theta)$ is supposed to be known. The only unknown quantity is the true value θ_0 of the parameter θ.

The next best thing to knowing the exact value θ_0 would be to have an upper and lower bound for θ_0. Thus we are led to try to determine two numbers $U(x_1, \cdots, x_n)$ and $L(x_1, \cdots, x_n)$ depending on n observations $x_1, \cdots, x_n$ on the random variable X such that $L(x_1, \cdots, x_n) \leqq \theta_0 \leqq U(x_1, \cdots, x_n)$. However this is only possible if we are willing to take a definite risk of obtaining incorrect limits. Thus we can fix a confidence coefficient α, $0 < \alpha < 1$, and then determine two expressions L and U—for simplicity we omit from now on to indicate that L and U depend on the observed sample values—in such a manner that $L \leqq \theta_0 \leqq U$ with probability α. In other words if we perform a great many experiments, compute each time L and U corresponding to the same confidence coefficient α, and state every time that the true parameter value θ_0 lies between L and U, we shall in the long run make correct statements $100\alpha\%$ of the time.

In the case when the form of the distribution function is known except for several unknown parameters confidence regions can be defined in a similar manner. However, as we have seen, if we assume only that the unknown cdf is continuous a finite number of parameters is no longer sufficient to specify the distribution completely. Now in order to know $F(x)$, we have to know its value for every x, $-\infty < x < +\infty$. Thus instead of looking for two numbers L and U as in the parametric case we should now look for two functions $L(x)$ and $U(x)$ defined for all x and then state that

$$L(x) \leqq F(x) \leqq U(x), \qquad -\infty < x < +\infty. \tag{1}$$

As in the parametric case we ought to indicate that both $L(x)$ and $U(x)$ depend on the observations $x_1, \cdots, x_n$, but for simplicity shall again omit to do so. As before, it is impossible to determine $L(x)$ and $U(x)$ in such a way that (1) is always true, but again we can fix a confidence coefficient α, $0 < \alpha < 1$, such that in the long run (1) is true $100\alpha\%$ of the time. We shall say that $L(x)$ and $U(x)$ determine a confidence band for the unknown cdf $F(x)$ corresponding to the confidence coefficient α, meaning that the band determined by the graphs

of $L(x)$ and $U(x)$ covers the graph of $F(x)$ completely with probability α.

As in the parametric case there are infinitely many ways of determining $L(x)$ and $U(x)$. Very little work has been done so far in finding confidence limits in the non-parametric case which could be termed *best.* However from the standpoint of facility of application one class of functions $L(x)$ and $U(x)$ seems to have definite advantages. Before we can describe this class of functions we have to define first what we mean by the sample cdf.

Let again $x_1, x_2, \cdots, x_n$ be a sample of size n from a population having cdf $F(x)$. For simplicity we shall assume that $x_1 \leqq x_2 \leqq \cdots \leqq x_n$. Since we are not interested in the order in which the sample was drawn this is no restriction. The sample or empirical cdf $F_n(x)$, as it is sometimes called, is now easily constructed. It is the step function which is equal to 0 for $x < x_1$ and equal to 1 for $x \geqq x_n$, while increasing by $1/n$ at each of the values x_i, $i = 1, 2, \cdots, n$. Thus we can write $F_n(x) = 1/n$ (the number of observations which are $\leqq x$). It can be shown that as $n \to \infty$ $F_n(x)$ converges stochastically to $F(x)$, or, in other words, that as n increases we can be almost certain that $F_n(x)$ approaches $F(x)$ more and more.

Determination of $L(x)$ and $U(x)$. The convergence of $F_n(x)$ to $F(x)$ which we have just stated suggests the following method of defining the lower and upper boundary of a confidence band for $F(x)$:

$$
(2)\qquad
\begin{aligned}
L(x) &= \begin{cases} F_n(x) - d & \text{if } \ F_n(x) - d > 0 \\ 0 & \text{otherwise} \end{cases} \\
U(x) &= \begin{cases} F_n(x) + d & \text{if } \ F_n(x) + d < 1 \\ 1 & \text{otherwise} \end{cases}
\end{aligned}
$$

where $d > 0$ is a constant determined in such a way that (1) is satisfied with probability α. Obviously, d is a function of α and n.

No formula is available to determine the value of d for any given α and n. However, Wald and Wolfowitz [1] have shown how to compute α when d and n are given. Though it would appear from (1) that α should also depend on $F(x)$, this is, fortunately, not the case. Thus a double entry table for α corresponding to values of d and n could be computed. From such a table the value of d which for a fixed n corresponds to a given value of α could be found by interpolation.

The following scheme shows the computation that has to be performed to find α. Compute $2n$ numbers, a_i and b_i, $i = 1, 2, \cdots, n$,

where

$$a_i = \begin{cases} \dfrac{i - nd}{n} & \\ 0 & \text{otherwise} \end{cases}$$

$$b_i = \begin{cases} \dfrac{i - 1 + nd}{n} & \text{if } \dfrac{i - 1 + nd}{n} < 1 \\ 1 & \text{otherwise.} \end{cases}$$

Define n functions $P_{k+1}(x)$, $k=0, 1, \cdots, n-1$, with the help of the recursion formula

$$P_0(x) \equiv 1, \qquad P_{k+1}(x) = \int_{a_{k+1}}^{u} P_k(t)dt$$

where

$$= \begin{cases} x & \text{if } x < b_{k+1} \\ b_{k+1} & \text{otherwise.} \end{cases}$$

Then

$$\alpha = n!P_n(1).$$

It is easily seen how the fact that we have to consider various upper limits at each application of the recursion formula makes the computation of α very cumbersome. However there is a very good approximation to α involving considerably less computational work. Set $\alpha = 2\bar{\alpha} - 1$, where $\bar{\alpha}$ is determined in the following way. Take the same a_i, $i=1, 2, \cdots, n$, as before. Define n functions $\bar{P}_{k+1}(x)$, $k=0, 1, \cdots, n-1$, with the help of the recursion formula

$$\bar{P}_0(x) \equiv 1, \qquad \bar{P}_{k+1}(x) = \int_{a_{k+1}}^{x} \bar{P}_k(t)dt.$$

Then

(3) $$\bar{\alpha} = n!\bar{P}_n(1).$$

Now only one integration is necessary at each application of the recursion formula. This procedure can be reduced to the evaluation of a determinant of order $n+1$ [1]. Above approximation of α with the help of $\bar{\alpha}$ is such that it increases our protection in the sense that the true probability of our confidence band covering $F(x)$ completely may be larger, but is never smaller than $2\bar{\alpha}-1$.

Discontinuous cdf's. As mentioned in the beginning we have assumed throughout that the unknown cdf is continuous. Under this assumption the probability that two sample values are equal is zero. In practice, however, due to limitations of measurement, it is quite possible that two or more sample values turn out to be equal. In such a case it may happen that for some sample value $x=x_i$, say, the upper limit $U(x)$ for $x<x_i$ is lower than the lower limit $L(x)$ for $x \geqq x_i$, making it impossible to draw a continuous cdf between $L(x)$ and $U(x)$. In such a case we shall be justified in asserting that the true distribution has a discontinuity at $x=x_i$.

The question arises what is the probability that our confidence band constructed as described above covers completely the true cdf if it should be discontinuous. We can no longer state that this probability is exactly α, but it has been shown that in this case we have

$$P\{L(x) \leqq F(x) \leqq U(x)\} \geqq \alpha$$

so that actually our protection is better than claimed.

Asymptotic Results. It is evident that it would be very desirable to have tables of the kind described earlier. In the meantime certain asymptotic results are available. If we let

$$\lambda = d\sqrt{n}. \tag{4}$$

Smirnoff [2] generalizing a result by Kolmogoroff has shown that

$$\lim_{n\to\infty} \alpha = 1 - 2\sum_{j=0}^{\infty} (-1)^{j-1} e^{-2m^2\lambda^2}. \tag{5}$$

Since (5) contains a very fast converging series, it is not difficult to compute λ corresponding to a given confidence coefficient α to any desired degree of accuracy. The corresponding value of d is easily determined from (4) to be

$$d = \frac{\lambda}{\sqrt{n}}\cdot \tag{6}$$

A short table of λ-values is given in Kolmogoroff [3]. These values are based on a table by Smirnoff [4].

Formula (4) may also be used in a different way. For a given investigation we may want to fix not only the confidence coefficient α but also the width of the confidence band we are going to obtain before-

hand. Then with the help of (4) we can determine the size n of a sample which will satisfy these requirements,

$$n = \frac{\lambda^2}{d^2}. \tag{7}$$

Thus if we should decide that we want a 95% confidence band such that the upper and lower bounds for $F(x)$ should not be more than .2 apart, we have to take $\lambda = 1.35$ and $d = .1$. Substituting in (7) we find that we shall have to use a sample containing at least 183 observations.

It is interesting to note that the width of the confidence band is inversely proportional to the square root of the sample size. This is equivalent to saying that the accuracy of our statements is directly proportional to the square root of the sample size.

One-Sided Limits. As in the parametric case when it may happen that we are only interested in an upper (or lower) limit for the unknown parameter, it may happen in the non-parametric case that we only need an upper (or lower) limit for the unknown cdf. Exactly the same method applies as in the two-sided case, except that now we have to substitute $\bar{\alpha}$ as computed by (3) for α, i.e., we can state with confidence coefficient $\bar{\alpha}$ that $F(x) \leqq U(x)$ (or $L(x) \leqq F(x)$), $-\infty < x < +\infty$.

An asymptotic approximation is also available. Smirnoff [2] has shown that

$$\lim_{n \to \infty} \bar{\alpha} = 1 - e^{-2\lambda^2} \tag{8}$$

where again $\lambda = d\sqrt{n}$ as in (4). (8) is very easily solved for λ. Indeed we get $\lambda = +\sqrt{-\frac{1}{2}\lg(1-\alpha)}$ where the logarithm is the natural or Naperian logarithm.

Goodness of Fit. It may be worth while pointing out that $L(x)$ and $U(x)$ can also be used to test the hypothesis that the unknown cdf $F(x) = F_0(x)$, where $F_0(x)$ is some given cdf. If we reject this hypothesis whenever $F_0(x)$ intersects either $L(x)$ or $U(x)$ or both, we shall be using a test with a critical region of size $1-\alpha$.

Confidence Bands Giving Confidence Intervals for $F(x)$ at a Specified Value x. The method which we have just discussed assures us a confidence band that with probability α covers the true cdf $F(x)$ in its entirety. It is not difficult to see, however, that for any given $x = x_0$, say, the probability that the corresponding interval $[L(x_0),\ U(x_0)]$

will contain the true value $F(x_0)$ is often considerably greater than α. It follows that if we are only interested in finding a band which for a given, though arbitrary value x_0 contains the true value $F(x_0)$ with probability α, we can use a narrower band, thus increasing the accuracy of our statements. Such a band is easily found. In fact, let $F(x_0)=p$. Then p can be considered as the unknown parameter of a binomial variate X defined by $P\{X \leqq x_0\}=p$ and $P\{X>x_0\}=1-p=q$. We have reduced the problem to a parametric one the solution of which is well known.

Using $F_n(x_0)$, where $F_n(x)$ is the sample cdf as before, as the sample estimate of p we can find two quantities $L'(x_0)$ and $U'(x_0)$ in such a way that $P\{L'(x_0) \leqq F(x_0) \leqq U'(x_0)\}=\alpha$. Now x_0 has been completely arbitrary. Letting it take all values from $-\infty$ to $+\infty$ we get two functions $L'(x)$ and $U'(x)$ determining a confidence band which satisfies our requirements. To distinguish between the two confidence bands we have obtained we shall refer to the one given by $L(x)$ and $U(x)$ as the type 1 and the one given by $L'(x)$ and $U'(x)$ as the type 2 band.

If the sample size n is sufficiently large so that we can use the normal approximation to the binomial distribution $L'(x)$ and $U'(x)$ are given by

$$(9)\qquad \begin{aligned} &\frac{n}{n+t^2}\left[F_n(x)+\frac{t^2}{2n}-t\sqrt{\frac{F_n(x)[1-F_n(x)]}{n}+\frac{t^2}{4n^2}}\right] \\ &\frac{n}{n+t^2}\left[F_n(x)+\frac{t^2}{2n}+t\sqrt{\frac{F_n(x)[1-F_n(x)]}{n}+\frac{t^2}{4n^2}}\right] \end{aligned}$$

respectively, as is shown, e.g., in Cramér [5], p. 514, Ex. 2, where t is the $100(1-\alpha)\%$ value of a normal deviate.

Let $w(x)=U'(x)-L'(x)$ be the width of the type 2 confidence band. From (9) we find

$$(10)\qquad w(x)=\frac{2nt}{n+t^2}\sqrt{\frac{F_n(x)[1-F_n(x)]}{n}+\frac{t^2}{4n^2}}\,.$$

Thus the width is no longer a constant as it was for the confidence band of type 1, but is now a function of x having its maximum value for those x for which $F_n(x)=1/2$ or is closest to 1/2.

It is instructive to compare this maximum width with $2d$, the width of the type 1 band, in an example. Let $\alpha=.95$ and $n=216$, the size of

the sample we shall consider later. We find $2d=2\lambda/\sqrt{n}=2.7/\sqrt{216}=.184$. For $F_n(x)=1/2$ (10) reduces to $t/\sqrt{n+t^2}$. From a table of normal deviations we find $t=1.96$. Substituting we get $w=.132$, which is quite an improvement over .184.

So far we have assumed that we were able to make use of the normal approximation to the binomial distribution. If this approximation is not accurate enough, we have to compute binomial confidence intervals as shown by Clopper and Pearson [6]. Now for $x_k \leqq x < x_{k+1}$, $k=0, 1, \cdots, n$, $x_0=-\infty$, $x_{n+1}=+\infty$ we have

$$L'(x) = \eta_k, \qquad U'(x) = \theta_k \tag{11}$$

where η_k is the lower, θ_k the upper binomial confidence limit corresponding to the observed frequency ratio $F_n(x_k)$. The original graphs of Clopper and Pearson together with others have recently been reproduced in Eisenhart [7], pp. 332–335. These graphs show η and θ as functions of the observed sample ratio for $n=10$, 15, 20, 30, 50, 100, 250, 1000 corresponding to $\alpha=.95$, .99 and for $n=5$, 10, 15, 25, 50, 100, 250, 1000 corresponding to $\alpha=.80$, .90. For other values of n η and θ have to be found by interpolation.

The exact values of η_k and θ_k corresponding to any n are given by

$$I_{\eta_k}(k, n - k + 1) = 1 - \alpha/2 \tag{12}$$

$$I_{1-\theta_k}(n - k, k + 1) = 1 - \alpha/2, \tag{13}$$

where $I_x(p, q)=\int_0^x t^{p-1}(1-t)^{q-1}dt/\int_0^1 t^{p-1}(1-t)^{q-1}dt$ is the incomplete beta function. By definition $\eta_0=0$, $\theta_n=1$. It is sufficient to solve either for the η's or the θ's since by (12) and (13)

$$\eta_k = 1 - \theta_{n-k}.$$

To find, e.g., η_k we can make use of the tables of percentage points of the incomplete beta function by Thompson [8], entering these tables with $\nu_1=2(n-k+1)$ and $\nu_2=2k$ on the page giving the $100(1-\alpha/2)$ percentage points.

When introducing the confidence band of type 2 we stated that we wanted a confidence band such that the probability that for any arbitrary x_0 $L'(x_0) \leqq F(x_0) \leqq U'(x_0)$ was α. This statement may have been somewhat misleading. We cannot make this probability exactly equal to α. For large n we used the normal approximation, thus committing a slight error, while for small n due to the discontinuous character of the binomial distribution exact confidence intervals do not exist, and we have to be satisfied with the statement that the confidence coefficient is $\geqq \alpha$.

It may be well to illustrate the use of the two types of confidence bands by some examples. An economist may want to analyze the income structure in a given community. If he is interested in the distribution of income over a specific range he will need a type 1 band, since he is looking for the joint occurrence of certain events. If, on the other hand, he only wants to state that at least $l\%$ and at most $u\%$ of the population earn no more than \$.... a year a type 2 band will give him the appropriate answer.

It may happen that we are not interested in an upper but only in a lower limit. A social worker who wants to prove the need for a new hospital in a certain section will want to state that at least such and such a percentage of the residents earn less than \$.... a year and thus cannot afford to go to a private hospital if the need should arise. The answer will be given by a type 2 band constructed with the help of one-sided confidence intervals. Then $U'(x)\equiv 1$, while $L'(x)=\eta_k$, where again η_k is the solution of (12), except that this time the right hand side should read $1-\alpha$.

Confidence Intervals for Quantiles. A confidence band of type 2 can also be used to construct confidence intervals for quantiles, i.e., confidence intervals for the value q_p for which $F(q_p)=p$, $0<p<1$. Such a confidence interval consists of all those values x for which $L'(x)<p<U'(x)$. These values of x are bounded by two observations, x_i and x_j, say, x_i being the smallest value for which $U'(x)>p$, x_j the smallest for which $L'(x)\geqq p$. If α is the confidence coefficient connected with our confidence band, we can then say that

$$x_i \leqq q_p \leqq x_j$$

is a confidence interval with confidence coefficient α for the unknown quantile q_p, i.e., in the long run confidence intervals chosen in this way will include the true value q_p $100\alpha\%$ of the time.

If a type 2 confidence band has been constructed x_i and x_j can easily be read off. However, we can find the two values also algebraically. To be exact, using the formulas for $L'(x)$ and $U'(x)$ we can find two integers i and j such that the corresponding observations x_i and x_j are the two observations we are looking for.

For large n the type 2 confidence band is given by (9). Let as usual $[y]$ stand for the largest integer $\leqq y$ and set

$$\begin{aligned} y_1 &= np - t\sqrt{np(1-p)}, \\ y_2 &= np + t\sqrt{np(1-p)}, \end{aligned} \tag{14}$$

where as in (9) t stands for the $100(1-\alpha)\%$ value of a normal deviate. Then i and j are found to be given by

$$(15) \qquad i = [y_1] + 1, \qquad = [y_2] + 1,$$

unless y_2 is an integer itself, in which case $j = y_2$.

Since we are only using an approximate confidence band, the solution given by (15) may sometimes lead to a confidence interval the true confidence coefficient of which is slightly $<\alpha$. Therefore, a safer, though very often unnecessarily wide confidence interval is given by $i = [y_1]$, $j = [y_2]+1$. It should also be remembered that if p (or $1-p$) is small the normal distribution is not well suited to serve as an approximation to the binomial distribution, even if n is relatively large.

If the type 2 confidence band is given by (11) i and j are found to be defined by the relations

$$\theta_{i-1} \leqq p, \qquad \theta_i > p,$$
$$\eta_{j-1} < p, \qquad \eta_j \geqq p,$$

where the η's and θ's are given by (12) and (13).

A case of special interest arises when $p = 1/2$. Then q is the median M of the unknown distribution. For this case Nair [9] has tabulated the values of i and j for $n \leqq 81$ corresponding to $\alpha = .95$ and $n \leqq 76$ corresponding to $\alpha = .99$.

We shall close with one specific application. In biological work it is often important to find a confidence interval for the median lethal dose of a given drug. This is the dose which would kill 50% of the animals of a given population. In the second experiment described in Bliss [10] the individual lethal doses of digitalis (cc. of tincture per kg. of life weight) of 216 test animals, in this case cats, were obtained. These data,[2] after adjustment for laboratory differences, can, with some caution,[3] be considered as a random sample of individual lethal doses.

The usual procedure for finding a confidence interval for the median dose is to take logarithms of the observations and assume that the log-doses are then normally distributed. If we are willing to make this assumption a confidence interval for the median log-dose can be found by a well known procedure which, however, involves considerable computation. The fact that the log-doses are approximately normally distributed has been observed in many experiments of this kind, but Mather [12], esp. p. 240, warns that for any new drug or new method of

---

[2] I want to thank Professor Bliss for putting these data at my disposal.

[3] See in this connection [11].

preparation the logarithmic transformation has to be justified anew. Then what shall we do in such a case if we have a sample of observations which is obviously not large enough to carry out a conclusive test of normality? An unjustified assumption of normality may lead to a serious mistake in our conclusions. However, we may use the more general procedure outlined in this section. The assumed continuity of the unknown distribution function is hardly a restriction. Besides, Scheffé and Tukey [13] have shown that the same procedure is applicable for all possible cdf's, if we only change the confidence statement slightly.

Let us now return to the data mentioned above and find a confidence interval for the median dose without assuming that by transforming to logarithms we can make use of normal theory. Substituting $p=1/2$ and $n=216$ in (14) we find at the 95% level

$$y_1 = 108 - .98\sqrt{216} = 93.6,$$

$$y_2 = 108 + .98\sqrt{216} = 122.4,$$

or by (15)

$$i = 94, \qquad j = 123.$$

All that remains to be done is to order the observations in increasing order of size and select the 94th and 123rd observations. Thus we find the confidence interval

$$.640 \leqq M \leqq .671.$$

It may be worth while emphasizing at this point the extreme ease with which all the non-parametric procedures we have encountered can be carried out.

A more important problem of biological assay than determining a confidence interval for the median dose is to find a confidence interval for the relative potency of a preparation of unknown potency compared with that of a standard preparation. As a matter of fact the data we used were originally of this kind. Again this problem is usually solved on the basis of normal theory, however, non-parametric methods are also available in this case. I hope to be able to come back to this problem at some later date.

One more remark before concluding—it is often stated that non-parametric methods lead to much wider confidence intervals than parametric methods, in particular the use of normal theory. This is

of course true, and parametric methods should be used whenever there is a sound basis for their application, unless the greater ease of non-parametric methods should outweigh the advantage that can be gained by the use of parametric methods. However, to assume that a sample has been drawn from a normal population for the sole reason of obtaining narrower confidence intervals is to defeat the purpose of statistics.

REFERENCES

[1] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, vol. 10 (1939), pp. 105–118.

[2] N. Smirnoff, "Sur les écarts de la courbe de distribution empirique," *Recueil Mathématique (Matematicheski Sbornik), New Series*, vol. 6 (1939), pp. 3–26.

[3] A. Kolmogoroff, "Confidence limits for an unknown distribution function," *Annals of Math. Stat.*, vol. 12 (1941), pp. 461–463.

[4] N. Smirnoff, "Table for estimating the goodness of fit of empirical distributions," *Annals of Math. Stat.*, vol. 19 (1948), pp. 279–281.

[5] H. Cramér, *Mathematical Methods of Statistics*, Princeton, 1946.

[6] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits in the case of the binomial," *Biometrika*, vol. 26 (1934), pp. 404–413.

[7] C. Eisenhart, M. W. Hastay, and W. A. Wallis, *Selected Techniques of Statistical Analysis*, New York, 1947.

[8] C. M. Thompson, "Tables of percentage points of the incomplete beta function," *Biometrika*, vol. 32 (1941), pp. 151–181.

[9] K. R. Nair, "Table of confidence intervals for the median in samples from any continuous population," *Sankhyā*, vol. 4 (1940), pp. 551–558.

[10] C. I. Bliss, "The U.S.P. collaborative cat assays for digitalis," *J. Am. Pharm. Ass'n*, vol. 33 (1944) pp. 225–245.

[11] C. I. Bliss and M. G. Allmark, "The digitalis cat assay in relation to rate of injection," *J. of Pharmacology and Exp. Therapeutics*, vol. 81 (1944), pp. 378–389.

[12] K. Mather, *Statistical Analysis in Biology*, New York, 1947.

[13] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics," *Annals of Math. Stat.*, vol. 16 (1945), pp. 187–192.

ON A METHOD OF ESTIMATING BIRTH AND DEATH RATES AND THE EXTENT OF REGISTRATION

C. Chandra Sekar
All-India Institute of Hygiene and Public Health, Calcutta
(on loan to the Population Division of the United Nations)

AND

W. Edwards Deming
Bureau of the Budget, Washington

A mathematical theory is presented which when applied to a comparison of the registrar's list of births and deaths with a list obtained in a house-to-house canvass, gives an estimate of the total number of events over an area in a specified period; also the extent of registration.

In the development of the theory, allowance is made for the fact that the chance of an event being missed on one list (registrar's list or the house-to-house canvass) may not be independent of its chance of being missed on the other list. Where there is likely to be lack of independence, a test is suggested and a method introduced to reduce the effect of dependence. This is done by subdividing the data into small homogeneous groups, such as might be formed by small areas, sex and age classes, domiciliary and institutional births; then by estimating the number of events in these groups separately and summing them for a total. The standard errors of the estimates are given.

The theory is applied to an enquiry that was conducted in February 1947 over an area known as the Singur Health Centre, near Calcutta, covering the years 1945 and 1946 separately, and it is found that the estimated total number of events for the area is usually greater when the estimate is built up by summing the totals for individual groups than when it is computed at once for the aggregated population. According to the theory this observation confirms positive dependence and indicates that the greater figure is nearer the truth.

The annual number of births and deaths in the Singur Health Centre (total population 64,000) is estimated subject to a standard error of from 1 to 3 per cent, and the registration is estimated to vary from about 40 to 70 per cent with a standard error of about 3 per cent. This enquiry provides basic ground work for the design of future surveys, and it is estimated that at a cost of Rs. 10,000 to Rs. 15,000 (3 rupees to the U.S. dollar) estimates of birth and death rates for an entire District in India with a population of one to two millions can be obtained with an overall standard error of about 5 per cent.

Purpose. The purpose here is to present a theory by which when vital registration is incomplete, an enquiry in the form of a house-to-house canvass may be used in conjunction with the registrar's list to estimate, *i.* the total number of births and deaths in an area over a specified period; *ii.* the birth and death rates; *iii.* the deficiencies in registration; and *iv.* the standard errors of all these estimates. The theory will first be presented, then applied to particular surveys in the Singur Health Centre.

Method of enquiry. The application of the theory which is to be developed requires a comparison of the entries on:

1. The registrar's list (referred to as R)
2. The result of a complete house-to-house canvass carried out by an interviewer (referred to as I) and the classification of the entries on these lists into the following four exhaustive groups:

C, the number of entries recorded in I and also in R (such entries, being found on both lists, are assumed to be correct without investigation).

N_1, entries recorded only in R but not in I, and after investigation found to be correct.

N_2, entries recorded only in I but not in R, and after investigation found to be correct.

X, entries recorded on one list or the other, but not both, and found after investigation to be incorrect.

This is a complete classification of the entries on the lists but not of the events. There will also be a number Y of events which are missed by both lists; this number will be estimated later by application of the theory.

Theory. Let N be the total number of events (births or deaths) in the specified period. Then an estimate $\hat{N}$ of N is furnished by the formula $\hat{N}=C+N_1+N_2+N_1N_2/C$ wherein N_1N_2/C is an estimate of the number of events Y missed by both R and I. This formula of estimation assumes that the chance of an event being missed on either list is independent of the chance of being missed on the other. A method is presented later on for investigating the validity of the assumption of independence and for introducing a modification where necessary.

It can be shown that: *i.* $\hat{N}$ is an unbiased estimate in the limit when N becomes large and the assumption just mentioned is valid; *ii.* the

maximum likelihood estimate is equal to $\hat{N}$ in the limit; *iii.* the standard error of N is $\sqrt{Nq_1q_2/p_1p_2}$. The last formula will be developed in the appendix. Here,

$$p_1 = \text{the chance of } R \text{ detecting an event}$$

$$p_2 = \text{the chance of } I \text{ detecting an event}$$

$$p_1 + q_1 = p_2 + q_2 = 1.$$

It follows that the better the performance in *either* R or I, the higher be p_1 or p_2, the smaller be q_1 or q_2 and the more precise be the estimate $\hat{N}$ of the total of events. It follows, moreover, that the precision of $\hat{N}$, expressed as a proportion (namely as a coefficient of variation), is $\sqrt{q_1q_2/Np_1p_2}$, wherefore if the theory be applied over an area large enough to contain a large number N of events, the total number N of events will be estimated with great relative precision.

The symbol p_1 is a measure of performance of the registrar, an estimate for which is $\hat{p}_1 = C/(C+N_2)$. This estimate $\hat{p}_1$ of p_1 is subject to a coefficient of variation of

$$\sqrt{\frac{q_1}{(C+N_2)p_1} \cdot \frac{N-C-N_2}{N-1}}$$

This error decreases as $C+N_2$ increases. For perfect performance on the part of the interviewer, $C+N_2=N$, and there is then no error in estimating the performance of the registrar.

The foregoing development is oversimplified. In practice there are some problems to take account of—incomplete investigation of the R lists; incomplete coverage of the population in the house-to-house canvass. Special types of events, like those occurring in institutions, are best taken care of as a separate group. Then again there is the problem of investigating the assumption mentioned above, and of measuring and correcting for the correlation between the chance of an event being missed by R and being missed by I. These points will be examined in the following paragraphs.

Effect of incomplete investigation of the registrar's lists. In the investigation of the R-lists there may be some entries left over unclassified by reason of incompleteness of entry, illegibility, or simple failure for any reason whatever on the part of the investigator to finish his job. So long as the correct entries amongst the unclassified entries on the R-list constitute unbiased samples from the two categories C and N_1 men-

tioned earlier, the omission of the unclassified entries from the calculations does not affect the estimation of N, the total number of events. The estimate of the extent of registration will be too low if the unclassified entries contain, as is likely, correct entries classifiable as C. If the unclassified entries are all counted as correct, contrary to fact, the calculations will lead to an overestimate of the extent of registration.

Effect of incompleteness of coverage of the population. As in every population enquiry, there will be some failures to elicit information from all the households. This will happen when some households in which an event took place have moved away temporarily or permanently, or when no responsible person can be found at home to give the information. So long as the events in the uninterviewed portion of households are included in the R-list to the same extent as those in the interviewed households, the estimation of N is unaffected. The calculation of N may therefore be little affected by incompleteness of coverage of the population.

The effect of institutional events. In rural areas the bulk of the births are domiciliary, but there are some small scattered hospitals drawing patients from a wide area, and a high proportion of the events that take place in them are for non-residents. The R-list may contain some or even all of the entries for these institutional events because the registrar is able to ascertain this information easily and accurately from the institutions. The interviewer, on the other hand, will, by the nature of a house-to-house canvass, fail to discover an institutional event concerning people who had no family connections in the area. Institutional events, as they are accurately ascertainable, are best handled as a separate block and not as a problem of estimation.

The effect of correlation between events missed on both lists. The first step is to define this correlation. The registrar and his co-workers will detect some events and miss others. The probability that the interviewer [I] will detect an event that was missed by R may be different from the probability that he will detect an event that was recorded by R. If these two probabilities are equal there is complete independence, but otherwise there is not, in which case the formula given above for the estimation of the total number of events will be incorrect. The extent of the error can be investigated. If as before,

p_1 = the probability of the registrar detecting an event

q_1 = the probability of the registrar missing it

then the probabilities in the 4 groups will be shown by the accompanying table, which defines four new probabilities, P_{21}, P_{22}, Q_{21}, Q_{22}. p and q are always complementary: $p_{21}+q_{21}=p_{22}+q_{22}=1$.

Group	Probability
C Detected by both	p_1p_{21}
N_1 Detected by registrar only	p_1q_{21}
N_2 Detected by interviewer only	q_1p_{22}
Y Missed by both	q_1q_{22}

If there is complete independence between the events missed by both R and I, then $p_{21}=p_{22}=p_2$, introduced previously, and $q_{21}=q_{22}=q_2$. When there is dependence the expected value of the estimate of the number of events Y missed by both R and I will be close to

$$\frac{Np_1q_{21}q_1p_{22}}{p_1p_{21}}$$

whereas the correct value is Nq_1q_{22}. The difference is

$$\frac{Np_1q_{21}q_1p_{22}}{p_1p_{21}} - Nq_1q_{22} = Nq_1\left(\frac{p_{22}}{p_{21}} - 1\right).$$

So if $p_{21}>p_{22}$, the total number of events is underestimated and if $p_{21}<p_{22}$, the converse. We surmise that $p_{21}>p_{22}$ is likely to be the case.

Similarly, in the case of dependance, the registrar's performance is estimated as $p_1p_{21}/(p_1p_{21}+q_1p_{22})$ instead of p_1, the difference being $(p_{21}-p_{22})p_1q_1/(p_1p_{21}+q_1p_{22})$. If $p_{21}>p_{22}$ the registrar's performance is overestimated and if $p_{21}<p_{22}$, the converse.

If

$$\begin{array}{ll} p_1 = .8 & q_1 = .2 \\ p_{21} = .6 & q_{21} = .4 \\ p_{22} = .4 & q_{22} = .6 \end{array}$$

the bias in the estimation of the total number of events will be

$$q_1\left(\frac{p_{22}}{p_{21}} - 1\right) = -.067 \text{ or } -6.7 \text{ per cent.}$$

This bias may be much more important than the standard error of an

estimate of the total number of events made under the assumption of zero correlation.

Method to reduce the effect of correlation. It is important to note that correlation signifies heterogeneity in the population for it implies that events that fail to be detected do not form a random sample from the whole population of events. This heterogeneity may arise only if there are differences in the reporting rates for different segments of the population, resulting in the group of failures being weighted disproportionately by the different segments.

It therefore follows that the correlation can be minimized by dividing the population into homogeneous groups and calculating the total number of events separately for each group; then by addition getting the grand total. In order to put this suggestion into practice, let us consider the difference between two estimates of the total number of events: *i.* by dividing the population into homogeneous groups and estimating the events in each group separately, then forming a grand total; *ii.* by treating the entire population as a unit. Let the population be comprised of k homogeneous groups, with N_i events in the i-th group $(i=1, 2, \cdots, k)$. Then let $p_1^{(i)}$ be the probability of the registrar detecting an event in the i-th group, and $p_2^{(i)}$ the corresponding probability for the interviewer. The expected value of the number of events missed by both in the i-th group is $N_i q_1^{(i)} q_2^{(i)}$ and for the entire population the total missed by both will be $\Sigma N_i q_1^{(i)} q_2^{(i)}$. As by definition there are only k homogeneous groups, this value will be estimated without bias when the groups are treated separately. But if the entire population of events were pooled, the expected value for the estimate of the number of events missed by both would be close to

$$\frac{\left[\sum N_i p_1^{(i)} q_2^{(i)}\right]\left[\sum N_i q_1^{(i)} p_2^{(i)}\right]}{\sum N_i p_1^{(i)} p_2^{(i)}}.$$

The difference in the two values will be

$$\frac{\left[\sum N_i p_1^{(i)} q_2^{(i)}\right]\left[\sum N_i q_1^{(i)} p_2^{(i)}\right]}{\sum N_i p_1^{(i)} p_2^{(i)}} - \sum N_i q_1^{(i)} q_2^{(i)} = -\frac{N^2 S_1 S_2 r}{\sum N_i p_1^{(i)} p_2^{(i)}}$$

where

$$S_1^2 = \frac{\sum N_i [p_1^{(i)} - \bar{p}_1]^2}{\sum N_i}$$

$$S_2^2 = \frac{\sum N_i [p_2^{(i)} - \bar{p}_2]^2}{\sum N_i}$$

$$N = \sum N_i \qquad \bar{p}_1 = \frac{\sum N_i p_1^{(i)}}{\sum N_i} \qquad \bar{p}_2 = \frac{\sum N_i p_2^{(i)}}{\sum N_i}$$

and

$$r = \frac{S_{12}}{S_1 S_2} = \frac{\sum N_i [p_1^{(i)} - \bar{p}_1][p_2^{(i)} - \bar{p}_2]}{S_1 S_2 \sum N_i}$$

is the correlation coefficient between $p_1^{(i)}$ and $p_2^{(i)}$, weighted by N_i, the number of events to which they have reference. If $r > 0$, then treating the entire population as a unit, we are led to an underestimation of the number of events missed by both parties and therefore an underestimation of the total number of events. This also results in an overestimate of the extent of registration. If this is the case, the population need be divided only to the stage when further division shows no increase in the total number of events. It should be possible by actual trial with some real data to decide whether (e.g., in computing number of deaths) 5-year age groups are a more effective subdivision than 10-year age groups; and whether infant deaths should be treated separately.

The enquiry in Singur Health Centre. The Singur Health Centre consists of four contiguous Union Boards,[1] viz., Singur, Balarambati, Bora, and Begumpur, situated in the Serampore sub-division of the Hooghly district. The village Singur which serves as the headquarters is only 21 miles away from Calcutta and is easily accessible by rail from Calcutta. The total area of the Centre is about 33 square miles and comprised of 68 villages with a total population of about 64,000 distributed over 12,000 families living in about 8,300 houses. As is usual in West Bengal, the villagers live close together in a compact block and wide fields separate such blocks. Since 1944 this area has formed the controlled practice field of the All India Institute of Hygiene and Public Health, Calcutta, for their experiment in Public Health Methodology.

Procedure for registration. The procedure for the registration of births and deaths in this area follows closely the method adopted in other parts of Bengal. The Chowkidar, i.e., the village headman, is the reporting agent and is required to submit periodically to the Sanitary Inspector,[2] who is the registrar of the area a list of births and deaths.

[1] The Bengal Province is divided into divisions, the divisions into districts, the districts into subdivisions, the subdivisions into thanas, and the thanas into Union Boards.

[2] A Sanitary Inspector is usually in charge of the health activities of a thana.

TABLE I

THE INVESTIGATORS' REPORT ON THE COMPARISON OF THE R AND I LISTS OF SINGUR HEALTH CENTRE

Event	Year	R: Registrars' lists						I: Interviewers' Lists
		Total number of events in the lists	Number Verified			No. non-verifiable, illegible, incomplete etc.	Number incorrect	Extra in interviewer's lists N_2
			Total	Common C found in the the interviewer's lists	Extra N_1 not found in Interviewer's lists			
Births listed as occurring in the village (excluding non-resident institutional)	1945	1,748	1,504	794	710	156	88	741
	1946	2,650	2,242	1,506	736	228	189	1,009
Deaths listed as occurring in the village (excluding non-resident institutional)	1945	1,356	1,083	350	733	190	83	372
	1946	1,052	866	439	427	117	69	421

With a view to improving the registration in this area, the voluntary services of a villager have been enlisted. He is not only expected to assist the *Chowkidar*, who may be illiterate, by making entries in the *Chowkidar's* register, but also to inform the registrar directly on all births and deaths in the village. The registrar also obtains a list of births, maternal and infant deaths as known to the Maternity and Child Welfare Department, and by co-ordinating the information from the three sources is expected to improve birth and death registration. For all practical purposes the voluntary agency began operating only from January 1946.

Method of enquiry. The enquiry in the Singur Centre covering 1945 and 1946 was started on the 17th February 1947. The field work lasted for eleven weeks. In this enquiry an interviewer called on every household to enumerate the resident population (separately as present and absent) and visitors with particulars of community, age, sex, and marital status, and to list all births and deaths which occurred in the village during 1945 and 1946, listing separately with relevant particulars those that occurred outside the Singur Health Centre. The lists so prepared are the I-list which, as was mentioned earlier, were compared with the registration books (the R-list). In the field-organization as actually employed, there were four investigators who worked at the comparisons and supervised the work of the 16 interviewers. The interviewers and the investigators were selected from the village population as it was thought that they would be able to obtain better cooperation than an outsider.

It should be emphasized that the comparison of the two lists is crucial. The establishment of the identity of two entries, one on one list and one on the other, sometimes requires extreme perseverance. In some cases the registrar's entry is by hearsay, and part of it may be wrong, and often much consultation is required. The interviewer's entry, however, is fortunately accompanied by a house-number or other means of identification by which the information may be verified if necessary.

Basic data obtained from the enquiry. Table I shows the results of the investigators' comparisons of the R and I-lists. As mentioned earlier, there are some problems arising from illegible and incomplete entries, the movements of the population and institutional births. The table gives some idea of the magnitude of these problems. For example the non-verifiable entries on the registrars' lists run to roughly 10% or more of their total entries. In view of their magnitude the assumption that the unverifiable entries are a representative sample of all entries, an assumption that will be made in the calculations, becomes all the more

important. The need of more careful registration in the future is apparent.

No separate account was maintained of the number of correctly registered events occurring in families that had migrated out of the village prior to the interviewers' survey. The assumption will be made that the registrars would have recorded this category to the same degree as for the non-migrants, but the number is small and under the conditions of the Indian village, this assumption is not important.

In this enquiry the non-resident institutional births and deaths are considered separately and excluded from the table, as indicated. Institutional facilities exist only in the Singur Union Board. The number of the institutional births to non-residents was about 8% in 1945 and 1946. The number of institutional deaths of non-residents was only about 3%.

Estimation of total births and deaths. In order to investigate the homogeneity of smaller groups comprising the whole, so as to arrive at the best estimate of the total number of events, calculations were carried out—

i. for the Centre as a whole (births and deaths)
ii. for each Union Board separately; then these figures were combined (births and deaths)
iii. for males and females separately for the Centre as a whole; then these figures were combined (deaths only)
iv. for age groups by sex for the Centre as a whole; then the figures were combined (deaths only)

In 1945 the total number of deaths as estimated by these four methods were 2234, 2238, 2245, and 2418 respectively each with a standard error of approximately 70. In 1946 the number of deaths as estimated by the four methods were 1,696, 1,684, 1,698 and 1,765, each with a standard error of approximately 40. The closeness of the first three estimates indicates that the chances of the registrar and the interviewer detecting an event did not vary to any marked extent between Union Boards and the sexes. The increase obtained by the fourth method clearly indicates that the chances of the interviewer and the registrar detecting a death may differ considerably with the age of the dead person. Positive correlation is indicated.

Higher percentages of deaths in the younger age-groups were missed by both R and I as compared with adult age groups. The proportion missed also show a tendency to increase in the more advanced age-groups. It would be interesting to ascertain whether the estimate could be increased still further by finer subdivision of age groups or by sub-

division in regard to other characteristics within each group, but no further analyses were conducted.

As for births, the total number estimated from the data of the entire Centre was 2908 for 1945 and 3744 for 1946. Separate estimation for the four Union Boards when totalled yields 2915 and 3775 for the same years. It is to be noted that while the latter figures are the higher of the two, the figure for 1945 is higher by only 1/7th of the standard error and the figure for 1946 is higher by a whole standard error.

The highest figure obtained by breaking the population into groups in various ways, and adding the estimated number of events, is to be accepted as nearest the true figure. The nonresident institutional events, which were left out of consideration may be added in to get the total number of events occurring in the area.

Estimation of rates and incompleteness of registration. For computing birth and death rates over an area, the population base is furnished by the house-to-house canvass. The total number of correct entries in the R-list judged against the total estimated number of events, measures the extent of registration. Tables II and III show the results obtained for rates and for completeness of registration.

TABLE II

BIRTH AND DEATH RATES IN 1945 AND 1946, SINGUR HEALTH CENTRE

	1945		1946	
	Rate	Standard error	Rate	Standard error
Birth rate per 1,000 population	46.1	0.8	59.8	1.0
Death rate per 1,000 population	37.7	1.2	27.5	0.7
Specific death rate (males)	36.4	1.6	27.3	1.0
Specific death rate (females)	39.2	2.1	27.8	1.0

TABLE III

PERCENTAGE OF BIRTH AND DEATH REGISTRATION DURING 1945 AND 1946

Union board	Birth registration		Death registration	
	1945	1946	1945	1946
Singur	60.4–67.9	70.9–77.1	38.1–46.9	42.0–49.1
Balarambati	51.5–55.8	53.3–57.8	45.8–55.9	50.8–58.0
Bora	53.1–61.3	56.0–66.0	54.9–66.5	52.6–63.4
Begumpur	47.4–50.3	61.3–64.7	42.6–46.4	44.9–48.1

Note (1) The range is due to non-verified entries on the *R*-list.
Note (2) The figures are subject to a standard error of about 3 per cent.

One comment may be made in regard to the birth rate for 1946, which appears to be very high. Possible explanation may be the improved economic situation after the famine of 1943, and demobilization. Another possible explanation is failure of the investigator to establish the identity of entries in the R and I lists, but if this were so, it should be more apparent for 1945, which it is not, as the birth rates for 1945 are much lower. An improbable explanation is that each Union Board is composed of extremely heterogeneous sections displaying negative correlation between the probabilities of detection of events by the Registrar and the interviewers.

Another comment should be made. The completeness of registration, recorded in Table III, is based on the number of *correct* entries in the R-list judged against the estimated total number of events. Official published rates in all countries are based on the total number of registrars' entries, correct plus incorrect, and the usual practice of inflating official rates to correct for incompleteness of registration yields spurious results: the rates are already partly inflated owing to incorrect entries. Proper inflation (correction of rates) is possible only by comparing the registration lists with the results of a population survey and making estimates of the total number of events and the proportion of incorrect entries in the registration lists.

The precision of estimated number of events. From the fact that the coefficient of variation of a total estimated number of events is $\sqrt{q_1q_2/Np_1p_2}$, it will be seen that the lower the efficiency of detection of an event on either the I or R-lists (p_1 or p_2), the greater the standard error of the total. In this enquiry, in spite of the fact that local people were hired and trained especially for this work, the efficiency of the interviewing was not of high order: only 67.2% of the births in 1946 and 52.8% in 1946 were detected by the interviewers. The corresponding percentages for deaths were 50.7 and 32.3. Methods of improving the performance of the interviewers must be sought, and it appears that the interval of time to be covered by the survey must not extend too far back.

It is highly important to bear in mind that regardless of the interviewers' performance, the method proposed here for estimating the total number of, N, events is not subject to bias,[3] but poor performance does increase the error of the estimate of N. It also increases the standard error of the estimate of the registrars' performance.

The coefficient of variation is also influenced by N. It is important to note that N in the formula refers to any total—not just a total over

[3] In making this statement the case of p_2 (or p_1) $=0$ is considered trivial and is excluded.

an area, but a total for any subgroup, such as an age or sex classes for which an estimate is prepared. For the area and sex classes that were used here, the standard errors of the estimated totals varied from 1 to 10%. Over a larger area, or over broader classes, the coefficients of variation would be reduced by the presence of the factor $\sqrt{N}$ in the denominator.

Costs. A few words regarding the cost of this particular enquiry may be helpful in planning future enquiries. The cost of the field-work, including salaries and overhead charges, amounted to Rs. 4,000. The cost of tabulation and analysis amounted to Rs. 1,500. The total cost was thus Rs. 5,500 or about $1\frac{1}{3}$ annas (2 U. S. cents) per capita in the area of enquiry. For various reasons (this being a pilot study and a complete listing of the population being desirable for other reasons), the entire population was covered without the introduction of sampling. In designing an enquiry for a larger area such as a province or even a district, sampling would be used.

For each area in the sample there can be calculated the total number of events and the rate: also the efficiency p_1 of the registrar's performance. For each sample-area, supposedly completely canvassed (no sub-sampling) there will be an error in estimating either the rate or the registrar's performance. The coefficient of variation in the rate will be the expression already given earlier, viz. $\sqrt{q_1 q_2 / N p_1 p_2}$. Likewise, the coefficient of variation of the estimate of p_1, the registrar's performance, is

$$\sqrt{\frac{q_1}{(C + N_2)p_1} \cdot \frac{N - C - N_2}{N - 1}}$$

Each symbol refers to the particular area covered. These errors are not erased by taking a complete canvass. (As a matter of fact, the particular enquiry described here was a complete canvass, yet subject to these errors.)

When sampling is introduced to study a whole District, the estimation of the total number of events, the rates, and the over-all efficiency of registration will be made by combining the data from a number of sample-areas. An additional error is then introduced for a District as a whole because of variability between the sample areas. The variability between the rates of the individual sample-areas may be much smaller than the variability between their total events, as it is usually difficult to define sample-areas of equal populations. It follows that usually a much smaller sample will provide a standard error of (e.g.) 4 per cent

in an over-all rate for a District than is required to provide the same precision in the total number of events.

The cost of attaining (e.g.) a 4% error of sampling will depend on the particular design of sample that is used; and the design in turn will, for greatest economy, depend on the density and distribution of the population, on the variability of the birth and death rates over the area for which estimates are to be prepared, on the costs of purchasing or preparing maps and lists by which the sampling procedure may be formulated, on the quality of personnel available to carry out the work, etc.

As a general principle, applicable to large populations, so far as the errors of sampling are concerned, the total number of cases (i.e., the total number of people, households, areas, or whatever unit constitutes the elements of sampling) to be included in the survey depends almost entirely on the precision of sampling that is desired in the estimation of the total number of events, or in the rate (whichever is the aim of the survey) and hardly at all *on the total number of inhabitants in the area to be covered.*[4]

In India, the birth and death rates should be estimated at least by the District (roughly 1 to 2 million inhabitants), and for smaller areas if funds would permit. Roughly speaking, to attain an over-all standard error of 5% (a reasonable aim for the present), the cost of a survey will run between Rs. 10,000 and 15,000 for a district.

Additional information provided. A survey of this type also provides[5] valuable ancillary information regarding other characteristics of the population such as size of family, age and sex distribution, marital status, occupation and industry, specific fertility rates, gross and net reproduction rates, and other information, but the list cannot be extended indefinitely because the interest of the field workers must not be dissipated too far from the main aims of the survey.

APPENDIX

THE STANDARD ERROR OF $\hat{N}$

An approximate value for the standard error of $\hat{N}$.

$$\hat{N} = \frac{(C + N_1)(C + N_2)}{-}$$

[4] It is presumed in this statement that the physical facilities for sampling (maps, lists, personnel, payment, etc.) are about the same over all parts of the area to be covered.

[5] As a matter of fact, the surveys reported have provided most of these additional items, and the cost mentioned includes them.

can be obtained by the application of the formula that the variance $Vf(x)$ of a function $f(x)$ of x is approximately given by

$$Vf(x) \doteq \left(\frac{\partial f}{\partial x}\right)_E^2 V(x)$$

where $(\ )_E$ denotes the substitution of the expected values for x appearing inside the bracket after differentiation, and $V(x)$ denotes the variance of x.

If $C+N_1$, $C+N_2$ and N are fixed, it is known that the expected value $E(C)$ and the variance $V(C)$ of C are given respectively by

$$E(C) = Np_1p_2$$

and

$$V(C) \doteq Np_1q_1p_2q_2$$

where

$$p_1 = \frac{C+N_1}{N}; \qquad p_2 = \frac{C+N_2}{N} \quad \text{and} \quad p_1 + q_1 = p_2 + q_2 = 1.$$

Under the same conditions, the variance $V(\widehat{N})$ of $\widehat{N}$ is

$$V(\widehat{N}) = (C+N_1)^2(C+N_2)^2 V\left(\frac{1}{C}\right)$$

which by the application of the formula given above reduces to

$$V(\widehat{N}) \doteq \frac{Nq_1q_2}{p_1p_2}.$$

The standard error of N is therefore

$$\sigma_{\widehat{N}} = \sqrt{\frac{Nq_1q_2}{p_1p_2}}$$

approximately.

EVALUATION OF PARAMETERS IN THE GOMPERTZ AND MAKEHAM EQUATIONS

J. F. Brennan

This paper describes a technique for determining the mortality characteristics of physical property through the use of accounting records of plant balances and yearly additions. Where time is not available for extensive actuarial research, the method produces results within tolerable limits of accuracy. Its limitations are pointed out.

The assembly of the statistics needed for an actuarial analysis of physical plant is a tedious and expensive task. Basic records are not always adequate for this purpose, a circumstance which poses a formidable research problem. The technique described herein was devised in an effort to bypass this obstacle. The application of the method requires only a money record of plant balances and of gross additions over a period of years.

Evidently the plant balance at any time is the summation of the survivors from the gross additions of all previous years. If, using this principle, we attempt a direct determination of the parameters for a Gompertz or Makeham equation, we immediately encounter the difficulty of making summations, because of the compound exponential property of these functions. We may, however, expand these functions into convergent power series and by disregarding terms beyond the second or third power, often find an acceptable solution. In the development of the theory in the following paragraphs, I take the case of the Gompertz equation and its power series, limited to the square term. The extension of the technique to the Makeham equation and to higher terms of the series expansion will be obvious.

The specific problem is to determine, on the basis of a set of data, the parameters of a Gompertz equation:

$$y = kg^{c^t}$$

where t = age and y = survivors at age t.

The data are given in the form:

$$P_1 = Y_{11}$$

$$P_2 = Y_{12} + Y_{21}$$

$$P_3 = Y_{13} + Y_{22} + Y_{31}$$

$$P_4 = Y_{14} + Y_{23} + Y_{32} + Y_{41}$$

where P_4 is the plant balance at the end of year four and Y_{23} is the

amount of plant remaining from the gross addition of year two, 3 years after installation, etc.

For simplicity in the development, there will be imposed the condition that at age zero, the survivor curve passes through point (0, 1), so that we show at each age the fraction surviving. Thus we write:

$$y = \frac{1}{g} g^{c^t}.$$

Expanding y in a Maclaurin series gives:

$$y = 1 + GCt + \frac{1}{2!} GC^2(G + 1)t^2 + \frac{1}{3!} GC^3(G^2 + 3G + 1)t^3 + \cdots .$$

where $G = \log_e g$ and $C = \log_e c$

Now if, as found in some applications, the values of G and C are such that over the range Y_{11} to Y_{1n} the sum of the cubic and higher order terms is small compared to the sum of the first and second order terms, then for that range the Gompertz equation may be approximated by

$$y = 1 - at - bt^2.$$

If coefficients a and b are determined from the data, the parameters of the Gompertz Equation are given by:

(1)
$$G = - \qquad a^2 + 2b,$$
$$C = -\frac{a}{G}.$$

To determine the values of a and b we proceed as follows: Let

(2) $$Y_{x,n-x.} = 1.00 - a(n - x) - b(n - x)^2$$

where $Y_{x,n-x.}$ = the survivors in year n, as a decimal part of the installations made in year x, (i.e. $(n-x)$ years after installation),
x = year of installation,
n = a particular year, subsequent to year x,
$(n-x)$ = age at year n,
a & b are constants to be determined.

If we sum up the survivors in year n out of the installations of all previous years, we have

(3) $$P_{n+1} = \sum^{n} A_x[1 - a(n - x) - b(n - x)^2]$$

where A_x = gross additions made in year x

P_{n+1} = capital in service (i.e. plant balance) at the end of the nth year, (beginning of year $n+1$).

Expanding (3) we obtain

$$P_{n+1} = \sum A - a[n\sum A - \sum Ax] - b[n^2\sum A - 2n\sum Ax + \sum Ax^2]$$

or

$$a[n\sum A - \sum Ax] + b[n^2\sum A - 2n\sum Ax + \sum Ax^2] = \sum A - P_{n+1}. \tag{4}$$

Now evidently values may be fixed for a and b by choosing any two different years n, and solving simultaneously the two resulting equations (4). Thus many different sets of values for a and b can be obtained by varying the selection of the two n's. A least squares solution based on a range of values of n is indicated.

For convenience let

$$U_n = [n\sum A - \sum Ax],$$
$$V_n = [n^2\sum A - 2n\sum Ax + \sum Ax^2],$$
$$Z_n = [\sum A - P_{n+1}].$$

Then Equation (4) may be written

$$Z_n = aU_n + bV_n. \tag{5}$$

From a set of equations (5) the best values of a and b ("best" in the lease squares sense) are found from the normal equations:

$$\begin{cases} a\sum U^2 + b\sum UV = \sum UZ, \\ a\sum UV + b\sum V^2 = \sum VZ, \end{cases} \tag{6}$$

where the summations are carried out over such a range of values of n as is appropriate to the approximation.

With the numerical values of a and b obtained by solving equations (6) the survivor equation (2) may be conveniently written

$$Y_t = 1 - at - bt^2 \tag{7}$$

where t = age = $(n-x)$, Y_t = survivors at age t as a decimal part of the additions made in year $0(Y_0=1.0)$

The values of G and C (and from them g and c) may now be obtained by substituting a and b in equations (1). The resulting Gompertz

curve and the parabola, equation (7), should then be plotted on the same graph. It may be found that the two curves exhibit important divergence in the higher years of the range employed in the solution. This indicates that the third order and higher terms of the power series should not have been neglected in Equation (2). To include higher order terms, however, would multiply the work of the solution tremendously, as will be obvious from the summations of equation (4).

A satisfactory way out of this difficulty consists in calculating a number of points on the curve, Equation (7), and to them fitting a Gompertz curve by King's method.[1] The range of points employed in this procedure should be approximately the same as that used in deriving constants a and b.

Chart I illustrates the processes heretofore discussed. The two curves shown thereon are calculated from the same actual book record. First the constants in the parabola were fixed by the method proposed herein. Two Gompertz equations were then found: one by substituting the parabolic constants in Equations (1); and the other by the logarithmic method of King. The latter method appears to give the better interpretation of the data, since, as seen on Chart I, it produces a curve coincident with the parabola over the range of the data employed.

How well the Gompertz equation, found by King's logarithmic method, defines the mortality characteristics of plant, may be judged from the results given in Table I, which shows how nearly the theoretical capital in service, calculated from the derived Gompertz equation, matches up with the actual book record. The example used is based on the same actual data as underlies Chart I and represents fixed capital investment in overhead electric conductors, to which yearly additions were erratic, i.e., not correlated with the amount of capital in service. The test is, therefore, especially significant. The standard error of estimate calculated from Table I is 7.7 or about $\frac{1}{4}$ of 1% of capital in service during the last 5 years of the table.

A further test of this technique was made on a group composed of thousands of homogeneous units of equipment (gas meters). Having previously, by the conventional actuarial process, derived a Gompertz survivor curve for this group, a sample accounting record of gross additions was assumed and a plant balance record calculated from the known Gompertz curve, covering a period of 30 years. The process described herein was then applied to the sample, King's method being used for the extrapolations. The results are depicted on Chart II.

[1] See "Textbook of the Institute of Actuaries" or Winfrey "Statistical Analyses of Industrial Property Retirements" Bulletin 125, Iowa Engineering Experiment Station, Ames, Iowa.

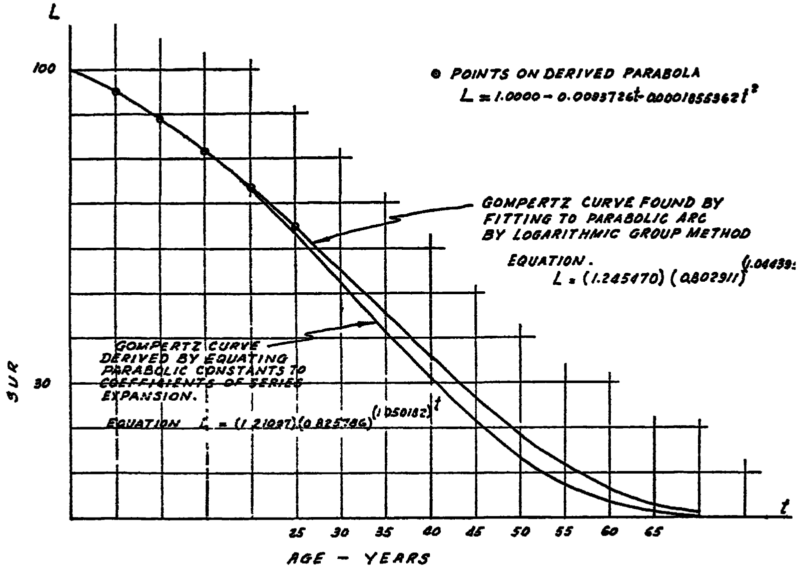


CHART I

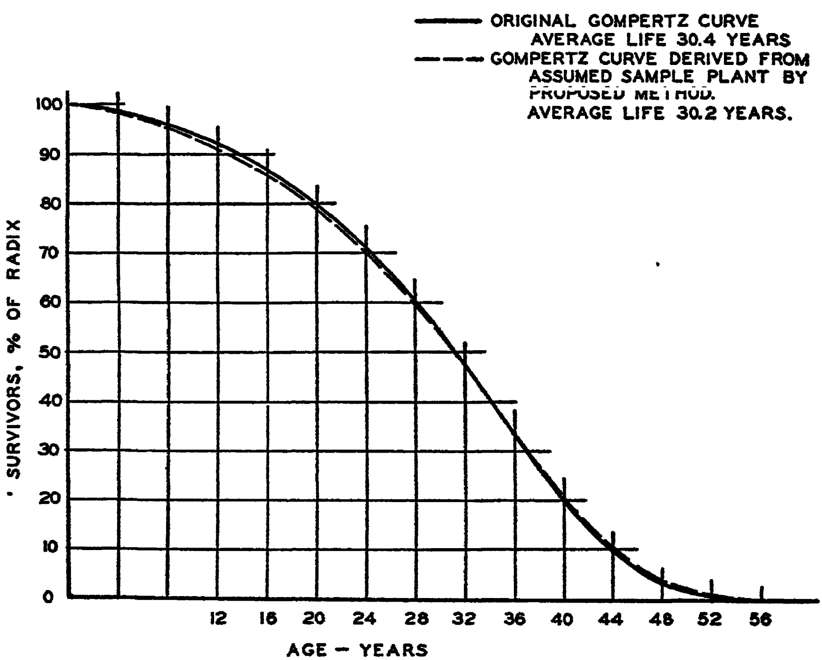


CHART II

TABLE I

TEST OF DERIVED SURVIVOR CURVE

Year	Capital in service at beginning of year (000 omitted)		Deviation (absolute value)
	Actual	Theoretical	
1935	$2,817	$2,840	23
1936	2,862	2,856	6
1937	2,916	2,910	6
1938	2,972	2,963	9
1939	3,097	3,098	1
1940	3,161	3,172	11
1941	3,223	3,229	6
1942	3,320	3,339	19
1943	3,391	3,390	1
1944	3,457	3,443	14

The true average life is 30.4 years, while that derived by the proposed method is 30.2 years, giving an error of approximately six tenths of one per cent. This is a tolerable deviation in this type of problem. No doubt greater precision could have been achieved by the use of a cubic (or higher degree) equation in lieu of the second degree parabola, but the extra labor is not justified in this case. Some precision was sacrificed in rounding out the data to units of one thousand dollars. Greater refinement is probably not warranted in the use of this process.

In applying this method, one may be confronted with a reliable statistical record covering a period of years, but may lack knowledge of the age distribution of the plant balance at the beginning of that period. Frequently in such cases a satisfactory solution can be made by estimating the average age of the beginning plant balance and treating it as a gross addition in the year corresponding to that age. Such a scheme has obvious frailty and should be employed only when no more logical process is possible.

It is conceivable that the derived parabola may have a maximum at an age significantly greater than zero. In some applications also, it may be found that the parabola is concave upward. Such results indicate a lack of stability in mortality ratios in the period (band) of years employed. This condition is entrained by shifting retirement policies, changed economic conditions, lack of replacement material or replacements and additions made by substitution of materials or equipment having inherently different life characteristics. Sometimes these faults will vanish with the selection of a different band of years (range of n) for solving Equations (6). If such conditions persist, however, the method breaks down.

ON THE "INFORMATION" LOST BY USING A t-TEST WHEN THE POPULATION VARIANCE IS KNOWN

JOHN E. WALSH
The RAND Corporation

This note calls attention to the use of the power function as a means of determining how much "information" is lost by using some other test in place of the most powerful test of a given hypothesis. As an example of the method, the case of using a t-test for the mean of a normal population with known variance is analyzed.

INTRODUCTION

If two significance tests of the same hypothesis should happen to have the same power function, these tests would furnish the same amount of "information" about the hypothesis tested in the sense of the Neyman-Pearson theory of testing hypotheses. Of course, it is hardly to be expected that two different significance tests will have *exactly* the same power function. In some cases, however, two significance tests may have very nearly the same power function. Then the two tests are said to furnish approximately the same amount of "information" concerning the hypothesis tested.

If there exists a most powerful test for a given hypothesis, "information" (in the sense of the Neyman-Pearson theory of testing hypotheses) will be lost by using some other test rather than this most powerful test. (For fixed sample size and significance level, a test is most powerful if the values of its power function are greater than or equal to those of the power function of any other test of the same hypothesis for the particular alternative considered.) It may happen, however, that the most powerful test (at same significance level) has approximately the same power function as the given test if the most powerful test is based on a smaller sample size; i.e., a most powerful test based on m sample values furnishes approximately the same amount of "information" as the given test using n sample values ($m \leq n$). Then it will be said that n-m sample values are "wasted" or "lost" by using the given test rather than the most powerful test. By convention, the value of m is allowed to assume non-integral values; the values of the power function of the most powerful test for non-integral m are found by interpolation from the power function values for integral m. This procedure furnishes an interpolated measure of the number of sample values "lost."

The above procedure could also be carried out in terms of operating characteristic functions rather than power functions. Since

$$\text{(Power Function)} = 1 - (OC \text{ Function}), \tag{1}$$

however, the same value of m is obtained.

The value of $100m/n\%$ is called the *power efficiency* of the given significance test. A discussion of power efficiency which contains an exact definition of when two power functions are to be considered equivalent (in the sense of furnishing the same amount of "information") is given in [1]. From (1), the definitions and remarks of [1] are equally applicable to the case in which OC functions are used instead of power functions.

As an example of application of the above method, let us consider a sample from a normal population with unknown mean and known variance. If it is desired to test the population mean with respect to a given constant value, the most powerful one-sided and symmetrical tests are based on the quantity

$$\frac{\text{(sample mean)} - \text{(given constant value)}}{\text{(population standard deviation)}}. \tag{2}$$

Thus, if the Student t-test is used instead of (2), "information" will be lost. This note presents an approximate expression for the number of sample values "lost" for the cases of one-sided and symmetrical t-tests.

The example analyzed has statistical interest in itself. Many statisticians have probably wondered how much information is lost when this situation occurs. One possible application of the result would be to help in deciding whether to use the t-test or test (2) with the population standard deviation estimated from past information. The final decision, of course, would also depend on the reliability of the estimate of the population standard deviation, the cost of taking observations, and perhaps other considerations. The complete formulation and analysis of this situation, however, is not considered to be a problem of this note.

Situations similar to the example analyzed here were investigated by Neyman in [2] and by Fisher in [3]. Fisher's results, however, are based on estimate rather than power function considerations.

2. *Results for example:* Let n sample values $x_1, \cdots, x_n$ be drawn from a normal population with unknown mean μ and known variance σ^2. Let us consider tests of whether μ differs from a given constant value μ_0 which are based on the t-statistic

$$t = \frac{(\bar{x} - \mu_0)\sqrt{n(n-1)}}{\sqrt{\sum (x_i - \bar{x})^2}}.$$

All the t-tests investigated in the note are of this type.

For one-sided t-tests at significance level α, or a symmetrical t-test at significance level 2α, it is found that approximately

$$\tfrac{1}{2}K_\alpha^2 n/(n-1)$$

sample values are "wasted" by using a t-test rather than the corresponding test of type (2). Here K_α is the standardized normal deviate exceeded with probability α; i.e., the function K_α is defined by

$$\frac{1}{\sqrt{2\pi}}\int_{K\alpha}^{\infty} e^{-x^2/2}dx = \alpha. \tag{3}$$

The above approximation to the number of sample values "lost" is reasonably accurate for $n \geqq 4$ if $\alpha = 5$ per cent, $n \geqq 5$ if $\alpha = 2.5$ per cent, $n \geqq 6$ if $\alpha = 1$ per cent, $n \geqq 7$ if $\alpha = 0.5$ per cent. The accuracy of the approximation increases as n increases.

If n is not too small, the above results can be roughly summarized by stating that $\frac{1}{2}K_\alpha^2$ sample values are "lost" by using a one-sided t-test at significance level α or a symmetrical t-test at significance level 2α. Table I contains values of $\frac{1}{2}K_\alpha^2$ for $\alpha = 5$ per cent, 2.5 per cent, 1 per cent, 0.5 per cent.

TABLE I

APPROX. NO. OF SAMPLE VALUES "WASTED" USING t-TEST WHEN VARIANCE KNOWN

Significance Level		Approximate No. of Sample Values "Wasted"
One-sided t- test	Symmetrical t-test	
5%	10%	1.4
2.5%	5%	2.0
1%	2%	2.7
0.5%	1%	3.3

3. *Derivations:* Let us consider the one-sided t-test of $\mu < \mu_0$ at significance level α and based on a sample of size n. Using a modification of the normal approximation given in [4], it is found that the power function values ϵ of the t-test are approximately determined by the relation

$$K_\epsilon = K_\alpha - \frac{(\mu_0 - \mu)}{\sigma/\sqrt{n}}\left[1 - K_\alpha^2/2(n-1)\right]^{1/2},$$

where the K_ξ function is defined by (3). This approximation to the power function is reasonably accurate for $n \geqq 4$ if $\alpha = 5$ per cent, $n \geqq 5$

if $\alpha = 2.5$ per cent, $n \geqq 6$ if $\alpha = 1$ per cent, $n \geqq 7$ if $\alpha = 0.5$ per cent. The accuracy of the approximation increases with n.

Now consider the one-sided type (2) test of $\mu < \mu_0$ at significance level α and based on a sample of size m. The power function values ϵ' of this test are exactly determined by the relation

$$K_{\epsilon}' = K_\alpha - \frac{(\mu_0 - \mu)}{}$$

Hence the two one-sided tests will have approximately the same power function if m is chosen so that

$$K_\alpha = \frac{(\mu_0 - \mu)}{\sigma/\sqrt{m}} = K_\alpha - \frac{(\mu_0 - \mu)}{\sigma/\sqrt{n}} [1 - K_\alpha^2/2(n - 1)]^{1/2},$$

i.e., so that

$$n - m = \tfrac{1}{2} K_\alpha^2 n/(n - 1).$$

Thus approximately $\frac{1}{2}K_\alpha^2\, n/(n-1)$ sample values are "wasted" if the one-sided t-test of $\mu < \mu_0$ at significance level α and based on n sample values is used rather than the corresponding type (2) test.

By symmetry, approximately $\frac{1}{2}K_\alpha^2\, n/(n-1)$ sample values are also "lost" by using the one-sided t-test of $\mu > \mu_0$ at significance level α and based on a sample of size n.

Now the power function of the symmetrical t-test of $\mu \neq \mu_0$ at significance level 2α and based on n sample values equals the sum of this power function of the one-sided t-test of $\mu < \mu_0$ with significance level α and sample size n plus the power function of the one-sided t-test of $\mu > \mu_0$ at significance level α and sample size n. Likewise the power function of the symmetrical type (2) test of $\mu \neq \mu_0$ at significance level 2α and based on a sample of size m equals the sum of the power functions of the two one-sided type (2) tests (of $\mu < \mu_0$ and $\mu > \mu_0$) at significance level α and sample size m. Thus approximately $\frac{1}{2}K_\alpha^2\, n/(n-1)$ sample values are "wasted" by using a symmetrical t-test of $\mu \neq \mu_0$ at significance level 2α and based on a sample of size n.

REFERENCES

[1] John E. Walsh, "Some significance tests for the median which are valid under very general conditions." Submitted for publication in *Annals of Math. Stat.* Abstracted in *Annals of Math. Stat.*, Vol. 18 (1947), pp. 610–611.

[2] J. Neyman, "Statistical Problems in Agricultural Experimentation," *Jour. Roy. Stat. Soc.*, Vol. 2 (1935).

[3] R. A. Fisher, *Design of Experiments*, Oliver and Boyd, 1942, p. 231.

[4] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), p. 376.

WESLEY CLAIR MITCHELL, 1874–1948

AN APPRECIATION

The death of Wesley Clair Mitchell brought to an end a lifetime of formative research, inspired scholarship, and earnest, continuous effort to apply scientific methods to social and economic problems. The end was untimely not merely because it always comes too soon for those few useful and lovable members of mankind of whom Dr. Mitchell was such an outstanding example. It was even more untimely because to the very end Dr. Mitchell retained the keenness of mind, the breadth of vision, the hospitality to new and pseudo-new ideas, and the kindliness to their often overconfident bearers—rare qualities even among scholars. To those of us who knew him and had the privilege of meeting him often, he seemed ageless and timeless. It is still difficult to realize that he is gone and will not be here to listen to our enthusiasms and complaints, to comment wisely and always with a charming humor upon some new quirk of the human mind, and to set before impatient younger generations further examples of broad scholarship and of respect for data and problems.

In these few notes it is perhaps most appropriate to stress Dr. Mitchell's work in statistics. Of the many who are familiar with his writings in recent decades only a few may realize how consistent was his interest and how continuous his research in the field of statistics. As a young graduate student at the University of Chicago at the end of the 1890's, his interest aroused by the monetary questions of the day, Dr. Mitchell was already contributing to the enrichment of quantitative knowledge by a series of articles on prices and by his work on the inflation experience during the Civil War—work that eventually resulted in two monumental treatises (published in 1903 and 1908). Upon completion of his graduate training at Chicago (with one year in Germany and Austria), Dr. Mitchell spent 1899–1900 at the Bureau of the Census, when Allyn Young and Walter F. Wilcox were there. This early combination of fruitful use of data in the study of economic problems with active interest in public agencies responsible for social and economic statistics set a precedent consistently followed throughout his lifetime. As work on currency and monetary problems gradually gave way to the broader studies on business cycles, Dr. Mitchell continued to maintain his active interest in and scrutiny of the basic data. There followed articles on the BLS index numbers of wages (*QJE*, 1911), on new banking measures (*JPE*, 1914), and on possible improvements in the statistical output of federal bureaus (Quart. Pub., ASA, 1915). From his

earliest productive years to the publication of that classical treatise on *Business Cycles* (1913) there was a continuous interplay of analysis, attempts at gathering and improving economic data, and efforts to raise the quality of basic information available to scholars and to the intelligent public at large.

This concern with bringing measurable facts to bear upon basic economic problems and with the need for critical scrutiny and evaluation of data made available by public agencies persisted throughout Dr. Mitchell's life. In 1915, barely two years after publications of *Business Cycles*, the BLS monograph on index numbers of wholesale prices appeared. This less well-known study, which was reprinted in 1921—an unusual distinction for a government report—is also a typical example of Dr. Mitchell's scholarship and approach—in the care with which the efforts of earlier scholars in the field are reviewed and utilized, in the breadth with which the problem is conceived, in the scrupulous attention paid to the characteristics of the available primary information, in the happy blend of insight and common sense with which the answers are provided and indeed the very questions formulated.

During the country's active participation in World War I, Dr. Mitchell served as Chief of the Price Section of the War Industries Board. Since he was always quite reticent about this period, one may surmise that it was not a happy one—for reasons which many scholars who passed through a similar experience in World War II can well understand. The pressure of urgent problems, the need for decisions made upon all too slim a factual basis, the tug and pull of various group and personal interests, hardly provided an atmosphere satisfactory to a scholar bent upon operating with wide and thoroughly weighed evidence. Yet several important and valuable results can reasonably be attributed to this experience. One was a clearer appreciation of the difficulties in the assembling of data and of research under government auspices—with some prescient suggestions for change, foreshadowing future reforms, made in Dr. Mitchell's Presidential Address to this Association in 1918 (see *JASA*, March 1919). Another was the series of monographs on the history of prices during the war, of which two volumes appeared under Dr. Mitchell's name. But perhaps the most important result of his war experience was the conviction that neither the university nor the government sufficed as loci of objective study of economic problems; and that a research institution, combining the continuity, theoretical interests, and the broad approach of the academic scholar with attention to quantitative data and the more realistic approach of government research, would plug a crucial gap and fill

a badly needed want. It was this conviction that provided the initial impetus to Dr. Mitchell and some of his wartime colleagues in the organization of the National Bureau of Economic Research in 1920.

Dr. Mitchell's own research since the early 1920's is closely associated with the National Bureau, of which he served as research director until 1946 and as an active member of the staff throughout and until the last. He headed the team that made the basic study of national income in this country in 1922 and set the pattern for work in the field that has grown apace ever since. It was his work on business cycles, in the increasingly broad conception of it as a pattern of change in the whole economy, that provided the central theme for all the work of the National Bureau through the almost three decades of its existence. It was his inspiration that held the National Bureau to standards that it endeavored to maintain; that attracted to it a group of people who combined theoretical interests with a zeal for established and testable evidence; and that kept the National Bureau from the temptation to take hard and fast positions on current and apparently pressing issues that were not warranted by the existing evidence. It was under the auspices of the National Bureau that Dr. Mitchell published his introduction to a new study of business cycles. *Business Cycles: The Problem and Its Setting* (1927) and the treatise on *Measuring Business Cycles* (jointly with Arthur F. Burns, in 1946). A report dealing with stable and variant characteristics of business cycles, now in preparation for publication by the National Bureau, engaged his attention during the last three years of his life.

Impressive as is this list of contributions during the last quarter century, it is incomplete in several respects. Dr. Mitchell was part author of many of the National Bureau publications, either as a direct contributor (to *Business Cycles and Unemployment*, *Recent Economic Changes*), or by his assistance rendered in review and criticism, or by the example and inspiration set by his own work. He served as guide and counselor to many other research organizations and projects—as chairman of the Research Committee on Social Trends (1929–33), as a member of the National Planning Board and of the National Resources Board (1933–35), as a member of the Social Science Research Council (since 1927)—to list but a few. The last service he rendered the government was as chairman of the technical committee set up by the National Labor Board to review the controversy over the BLS cost of living index (1943–44). And those who knew him were all too aware of how much of his time and effort was spent in counsel and guidance, kindly and modestly extended to all scholars, young and old, who were seeking it in increasing numbers.

The enriching influence of this lifetime of scholarship on research in the social sciences, on the teaching of economics and related subjects, and on public policy is well recognized and hardly requires demonstration. The ever widening use of statistical data and tools in the analysis of economic problems, the emphasis on the concrete institutional and historical framework within which societies live and function, the more scrupulous distinction between a recorded observation and plausible assumption, are all comparatively recent trends in the economic and social disciplines in this country and elsewhere. To this quickening of the searching spirit in the study of society, Dr. Mitchell's own investigations and those directed by him, were a major impetus. It should also be noted, for the benefit of those who are concerned with direct and practical utility, that the return to society from such effort is far greater than may appear on the surface. Its value to society is not only the obvious one of making possible more intelligent solutions of social and economic problems because more is known about the functioning of the economy—a clear illustration of this was provided in the work of our economic agencies in World War II. Its even greater, if less obvious, value lies in the spread of the spirit of inquiry and of the respect for facts, which impose desirable limits on the "mutable Minds, Opinions, Appetites and Passions of particular Men."

Yet the study of society through the use of statistical and other testable evidence is far from an easy task. Those who have wrestled with the complexity and variability of observable economic life and with the imperfect and treacherous data available on social phenomena, know the courage, patience, and sheer moral stamina required in this struggle and the unusual capacities for organization, analysis, and synthesis needed to bring order out of chaos. It would be useless, and perhaps impertinent, to inquire by what turn of the wheel that determines heredity and environment was Dr. Mitchell endowed so richly with all these qualities. But it is important to indicate, as well as one can, the leading ideas and the broad attitudes that assisted him throughout his lifetime.

These ideas or attitudes may be briefly stated under three heads. First was the conviction that the human mind is infinitely productive of hypotheses or models and that, from its rich endowment, it proceeds to originate them in profusion—regardless of the extent to which they are anchored in testable evidence. This conception led Dr. Mitchell to approach the products of the human mind with both respect and caution—respect for the rich insight and small modicum of experience that they may embody, and caution in accepting the wide interpretation and inference that are almost inevitably attached to the products.

There is a revealing discussion of this attitude in Dr. Mitchell's letter to J. M. Clark (see *Methods in Social Science*, edited by Stuart A. Rice, Chicago, 1931, pp. 674–80). It was an attitude particularly helpful in the field of social study in which group interests and passions tend unconsciously to color the hypotheses or models originated with such ease and with such a claim of finality.

Second, and equally important, was the dominant notion of interrelation in space and continuity in time as the basic characteristics of social and economic life. It is significant that the whole line of evolution in Dr. Mitchell's work is from prices (under the specific angle of inflation), to business cycles, and to the study of economic change at large. But this semblance of evolution is deceptive, since in his early investigations Dr. Mitchell already fully recognized that the study of any one part is in effect the study of the whole from a particular angle. This basic idea, in combination with the critical attitude toward man's theorizing, resulted in a natural emphasis on testable evidence and on an approach that, however meticulous with regard to the parts, never lost sight of the whole. The first attitude made Dr. Mitchell an empiricist; the second made him a synthesizer in the best sense of the word. The first helped him to resist the temptation to escape into the quiet haven of imaginative models and 'caeteris paribus'es; the second helped him to avoid refuge in the details of empirical work and kept him from indulging in the perfectionist's delight of whittling at minutiae until the Greek Kalends.

But there was a third and perhaps most basic idea—that there was some order in the ceaseless change and variance of economic phenomena; and that the patient building up of testable quantitative data, accompanied by the cautious and critical use of theories as hypotheses, might reveal the invariant elements. It is this idea that illuminated Dr. Mitchell's work with a steady glow, that served as a powerful magnet around which the detailed findings in his treatises arranged themselves in a comprehensible pattern. And it is the quest for this underlying order that provided the powerful drive in this long life of search and research—in the belief that as the pattern is gradually revealed and its concrete manifestations recognized, it will be accepted by human intelligence as the basis for action on social and economic problems.

Wesley Mitchell would have been the first to protest against such analysis in what he would consider grandiloquent terms: he was a modest and humble man—with a humility that, like all genuine humility, verged on pride. And I am writing these lines reluctantly. My only

justification is that it is important to realize what guiding ideas were helpful in a lifetime of fruitful and fundamental scholarship. It is important to recognize how strong today is the temptation to withdraw into the security of imaginary models, only distantly relevant to historical reality—regardless of how mathematically elaborate such models may be. It is important to see how ever present is the opposite temptation—to elaborate and check details without concern as to their place in the broader framework. And while these two lines of intellectual pursuit are moderately useful in the development of science, the third direction, equally tempting—to give up hope of finding any intellectual order and to resolve the problem by withdrawal into esthetic fancy or intellectual cynicism—can be of negative value alone. Dr. Mitchell's life is an inspiring demonstration of how effectively such temptations can be combatted, and how the spirit of objective inquiry can yield rich results in the study of human society.

To those who knew him well and to those who knew him slightly the passing of Wesley Mitchell is a great and numbing loss. These notes can give no idea of his personality, the quiet and often playful wit of his conversation, the genuineness of his moral seriousness, the consistent drive of his interests, the warmhearted attitude to and the broad tolerance of fellow men. His contributions to our knowledge of social phenomena, to the data on the basis of which more intelligent policy decisions are possible, to the training of a large group of scholars in the field, will stand; and, one may hope, will provide a foundation upon which work will be carried forward. But the personal loss is irretrievable, and we are all the poorer for it.

SIMON KUZNETS

BOOK REVIEWS

Edited by

OSCAR KRISEN BUROS

Rutgers University

Fraction-Defective Charts for Quality Control. *British Standards Institution.* British Standard 1313: 1947. London S.W. 1: British Standards Institution (28 Victoria St.), 1947. Pp. 40. 6s. Paper. (New York 17: American Standards Association [70 East 45th St.] $2.25.)

REVIEW BY ALBERT H. BOWKER

Assistant Professor of Statistics, Stanford University

Stanford, California

As THE title implies, this standard describes control charts for fraction defective and is a revision and extension of a small part of British Standard 600R:1942 *Quality Control Charts* reviewed by Harold A. Freeman in this JOURNAL (Sept. 1945, p. 386). The previous standard discusses, in addition to fraction defective, charts for mean, standard deviation, and several other statistics computed from continuous measurements. Further, it utilizes more technical knowledge of statistics than the present pamphlet, which emphasizes the applied side.

The limitation of the subject matter to the fraction defective chart, the omission of a discussion of the statistical principles behind the chart, and the internal organization of the pamphlet are designed to facilitate initial application of the method. The first section, entitled "The First Control Chart," suggests the application of a simple chart in a rigidly prescribed way. It recommends the selection of a product containing about 7% or 8% defective, using samples of twenty, sampling about 5% of the product, basing the process average on 25 samples, and using an upper limit exceeded with probability .005. Later sections discuss possible variation in amount of sampling, size of sampling, sampling interval, and probability limit. Other types of charts are discussed, including control charts based on a given standard rather than on an empirically-determined process average; two-way control charts, in which a separate control chart is kept for the per cent of items which exceed the upper limit and those which fall below the lower limit of an allowable range; and compressed limit charts for use when the fraction defective is very small. In this latter case, the control chart is based on the number of defectives outside gauge limits more stringent than the specification limits.

The discussion of the first control chart assumes that the product to which the chart is applied is already in control and discusses appropriate action when an occasional point falls on or beyond the control limit. A considerable number of cases have been reported in American literature in which the initial application of control charts leads to the discovery that the process

is badly out of control, with, in some cases, a majority of the points falling outside the control limits. Statistical control is apparently achieved only after a lengthy study and modification of the process. A description of this phenomenon might keep initial users from becoming discouraged if their individual sample qualities fail to cluster neatly around the process average, as in the examples provided here.

The discussion of control limits based on a given standard assumes the specification of a maximum permissible average level for defectives. Control limits are found by treating the specified percentage defective as the process average. It is clear that, if articles are produced with quality equal to this specified average, only an occasional point will be outside the control limits. Indeed, the presented quality would have to be a great deal worse than this maximum permissible average level before we could state with high probability that an out-of-control point would be obtained. Thus, the terminology "maximum permissible" is somewhat confusing.

In all the examples in the pamphlet, only upper limits on fraction defective are included. Customary American practice is to use both upper and lower limits. However, the reader is advised to investigate the cause of a consistent run below the process average. Another departure from common American practice is the use of probability limits, as opposed to 2σ or 3σ limits. Further, the expected number of defects per sample is less than some American authorities recommend or imply.

The pamphlet is well done and has several desirable features. The directions for setting up charts are very clear; the use of symbols has been almost entirely avoided; the instructions are reproduced conveniently in sample instruction sheets at the end of the pamphlet; there are several examples with practical advice; and there are references for further study.

The reviewer agrees with the review cited in the first paragraph, which concludes: "It is possible that this particular job has now been done well enough. The reviewer would welcome a pamphlet on the statistical theory on which quality control rests for this is not quite as obvious and ironclad as these excellent applied publications make it out to be."

Principles of Counting and Probability. *J. C. Abbott* (Associate Professor of Mathematics), and *T. J. Benac* (Associate Professor of Mathematics). (United States Naval Academy, Annapolis, Md.) Annapolis, Md.: U.S. Naval Institute, 1947. Pp. iii, 40. Paper. $1.00.

Review by Herbert Solomon
Air Intelligence Specialist, Air Intelligence Division
Headquarters, United States Air Force,
Washington 25, D.C.

This booklet is intended primarily for students in naval science, particularly naval gunnery. However, the illustrations and exercises are directly analogous to those encountered in aircraft gunnery and bombing procedures.

The booklet is divided into two chapters: I, "Principles of Counting" and II, "Probability." Answers are given at the end of the text to the 282 problems posed at intervals in the booklet. An attempt at every fourth exercise revealed no irregularities in the published answers.

Chapter I has some faint rumblings of set theory but mainly presents, in a very brief manner, material on permutations and combinations which can be found in any of the standard textbooks. Chapter II presents some of the fundamental theorems of probability in a simple non-rigorous manner which is, no doubt, exactly the intention of the authors. This lack of rigor will of course present no difficulties in working out the exercises or understanding the illustrations. The only probability distribution discussed is the binomial distribution and very little of its characteristics are studied. The multinomial distribution is left to the exercises. A noticeable omission is the normal distribution which plays a very important part in fire control studies in military and naval science.

As mentioned before, this booklet is designed for a rather special group but except for its illustrations and exercises the material contained can also be found in many easily accessible books on algebra and mathematical probability.

Business Cycles and Forecasting, Third Edition. *Elmer Clark Bratt* (Professor of Economics, Lehigh University, Bethlehem, Pa.). Chicago 4, Ill.: Richard D. Irwin, Inc. (332 South Michigan Ave.), 1948. Pp. xii, 585. $6.00.

REVIEW BY WILLIAM A. SPURR
Professor of Business Statistics, Stanford University
Stanford, California

THIS book is the first satisfactory general survey work published in its field since the war. Professor Bratt has condensed and considerably revised his last edition of 1940, which had suffered the rapid obsolescence typical of this period.

Like its predecessors, this edition begins with a brief treatment of seasonality and long term trends, and then covers the whole gamut of business cycles—their measurement, causes, theories, history, barometers, projection methods and proposals for stabilization. A postwar book of such scope has very real value for both the student and the business man.

The topic which has required the most complete revision since 1940 is that of business indicators or "barometers" (Chaps. 15–17). The concept of gross national product is now given central importance (at the expense of other general business indicators), and a method of projecting its components is offered later as the core of a five-step "effective program for business-cycle forecasting" (pp. 437–443). An actual case illustration, however, is needed to clarify this very promising method.

Another "one of the distinctive features of this third edition" which "adds leverage to the analysis" (p. v), is the treatment of secondary trends sepa-

rately from secular trends. The concept of the secondary trend (which is related to the long cycles or intermediate trends described so variously by Wardwell, Kondratieff, Juglar, Kitchin, Silberling and others), however, remains a shadowy one, as it does in Burns and Mitchell's *Measuring Business Cycles* (New York: National Bureau of Economic Research, 1946, Chap 11). The results are therefore inconclusive. Bratt properly points out that "regularity of recurrence of the secondary trend [is] a completely undemonstrated conclusion" (p. 71)—thereby avoiding a basic fallacy of Dewey and Dakin (*Cycles*. New York: Henry Holt and Co., 1947)—so that "no consideration can be given to the forecast of the secondary trend" (p. 77). Later, though, he seems unduly pessimistic in concluding that therefore *secular* trends may not be projected (p. 77).

The treatment of business forecasting (Chap. 18) is considerably condensed from that of the revised edition (Chaps. 21–22). One misses a detailed discussion of leads and lags, specific historical analogy and several other traditional forecasting procedures that may still have some validity today. Furthermore, the pessimistic statement (p. 420) that "we have practically no basis whatsoever for forecasting originating causes" (such as acts of government and wars) seems to be countered by the partial success of such Washington forecasters as Cherne and Kiplinger and by Bratt's own section on "Measurable Effects of Originating Causes" (pp. 401–402).

Other parts of the book are revised less radically. Seasonality, secular trends (including growth curves) and several methods of analyzing time series are presented in readable, nontechnical fashion. A preliminary chapter on "Concepts of Balance" in the revised edition has properly been omitted.

The detailed treatment of factors responsible for the cyclical nature of business (Chaps. 5–6), remains "the central part of the analysis" (rev. ed., p. vi), and provides a valuable analytical description of the course of a typical cycle. Other sections of particular interest are those on the distinction between difference and summation series (pp. 90–92) and on measures of business confidence (pp. 416–418).

The eclectic survey of business cycle theories (Chaps. 7–9) has been condensed and simplified since the 1940 edition, but is basically unchanged. The treatment here is perhaps not quite as lucid as in Estey's *Business Cycles* (New York: Prencice-Hall, Inc. 1941) but is more comprehensive.

The history of business cycles has been brought up to date through the end of 1946 (Chap. 14). The concluding section on parallels between World Wars I and II (pp. 350–352) is a provocative one which would justify more penetrating and detailed treatment.

The final section on proposals for stabilizing business cycles, has been expanded in line with the growing importance of this problem. The recent work of the President's Council of Economic Advisers is included.

The chief virtue of the book as a whole is its broad, impartial survey of all the main aspects of business cycles, reflecting the author's accumulated experience in preparing three editions of this work. This reviewer fully sub-

scribes to Bratt's basic policy of separating time series into secular, seasonal and cyclical-random elements, and his use of the analytical approach to business cycles developed by Wesley C. Mitchell in his 1913 and 1927 classics. The third edition is thoroughly modernized and is well organized. The faults are minor. While it is sometimes obscure or superficial, it provides many references to primary sources as guides for more intensive study. The condensation in general is an improvement, though the printer's crowding of more words per page makes for slightly more difficult reading than before. This book is recommended both as a text and as a general reference work.

Theory of Experimental Inference. *C. West Churchman* (Associate Professor of Philosophy, Wayne University, Detroit 1, Mich.). New York 11: Macmillan Co. (60 Fifth Ave.), 1948. Pp. xi, 292. $4.25. (London W.C. 2: Macmillan & Co., Ltd. [10 St. Martin's St., Leicester Sq.] 21*s*.)

REVIEW BY JOHN W. TUKEY

Assistant Professor of Mathematics, Princeton University, Princeton, N. J.
Member of Technical Staff, Bell Telephone Laboratories, Murray Hill, N. J.

THE author has tried to write a challenging book—and has succeeded. By mixing modern statistical inference and classical philosophy he has written a book which could serve to introduce statisticians to philosophy and philosophers to modern statistical inference. The book is a far from perfect tool for either job, but it will, in this reviewer's opinion, have substantial influence both on philosophers and on statisticians. It is, however, the meaning of the book to statisticians interested in the foundations of their subject which ts the chief topic of this review.

The first three chapters are devoted to a discussion from the point of view of formal science and philosophy of modern statistical inference, as examplified by the Neyman-Pearson theory, and a brief discussion of its relations with scientific method. The next chapter outlines a formal classification of systems of philosophy, according to their views on knowledge, into rationalism, naive empiricism, statistical empiricism, criticism, relativism and, finally, experimentalism. The first five schools are each the subject of a chapter, while experimentalism, founded by E. A. Singer and supported by the author, is discussed in four chapters. The book concludes with three chapters relating inference with social groups, social purposes, and a proposed science of ethics,

The most striking point to the statistician who is concerned with the foundation of his subject and who believes that the millennium is still far away is the complete acceptance by the author of a definite methodology of statistical inference as *the* methodology. The reviewer feels that the present methodology of statistical inference has been significantly biassed by desires for (i) analytic manageability, (ii) mathematical simplicity, and (iii) unwarranted uniqueness. The philosopher and the statistician now need to collaborate in working toward the ideal basis of statistical methodology.

Another striking point is the insistence of the author on security. He ad-

mits that complete freedom from risk is impossible in practice, but he insists that the possibility of reducing the risk to an arbitrary small value is essential. Again: "And he *must* be sure about these things, or else he would find it impossible to act efficiently; he cannot even entertain the notion that there are risks involved in his decisions, for if such doubts creep in, he finds it impossible to act quickly and efficiently" (p. 236). This sentence is supposed to refer to every day decisions, but it expresses the authors' general philosophy.

Finally: ". . . we must have criteria of the most efficient methods of solving problems before we can give responses to any questions" (p. 243). This statement is the more surprising when we recall that a "response" to the author is only an approximate step toward an answer.

The author's discussion of presuppositions and their importance in statistical inference (p. 12) should be read by all statisticians interested in methodology.

The author holds that: "In a sense, the problem of the best 'design' of an experiment is exactly the problem of the philosophy of science . . ." (p. 21). This is closely related to his quotation from R. A. Fisher: "The more thorough the design of the experiment, the more meaningful is the question asked" (p. 208). The author and the reviewer are in agreement as to the validity and importance of these quotations, but we draw differing conclusions. The author concludes that design of experiment transcends statistics, while the reviewer concludes that the philosophy of science is a part of statistics, since he defines statistics as: "The science, the art, the philosophy, and the technique of drawing conclusions from the particular to the general." Leaving this difference aside, it follows that any adequate account of the design of experiment must include serious attention to the philosophy of science. The author is led to the following strong statement: "But there should be realization in statisticians' minds that they have pushed their basic problem beyond the field of formal statistics when they attempt to set down the criteria of best test. The danger of not realizing this point lies in the possible action that will result when a formally defined criterion of best is taken to satisfy nonformal demands of the science of value" (p. 283).

The author's solution of the philosophical problem posed by randomness is: "*We would not be able to find randomness in our observations had we not first put it there in some form*" (p. 124). This is consistent with his desire for security and his faith in a future physics without indeterminacy (pp. 77 and 231–233).

Since the author's philosophy does not provide explicitly for the criticism of statistical presuppositions, it is not particularly surprising that, on page 284, he holds that tolerance limits can be set with equal validity from small as from large samples. For this would be correct if the usual presuppositions were correct.

At a more philosophical level, the author concludes that science, in any sense, can only exist when nature is regular—"That is, the meaning of an observation presupposes a principle of regularity in nature" (p. 128). This

position must, it seems to the reviewer, be accepted. But the author goes on to insist that:—"The reason that the relative frequencies must approach some limiting value is that the question of probability is otherwise meaningless; one is 'guaranteed' that they do by the natural image which is presupposed in all experimental problems" (p. 203)—and to insist that: "The fundamental postulate of experimentalism, therefore, is the following: *There exists a formalization of nature, such that stochastic limits exist for certain sequences of mathematical functions of the observations which are pertinent to a given question of fact*" (p. 178). This seems to the reviewer an overstrong and unwise requirement of *security*. The pertinence of the observations is, according to the author, to be settled by "formal" methods: ". . . the justification for assuming that a certain set of actions produces pertinent observations depends upon theoretical (formal) considerations on the part of the experimenter. These considerations must be presupposed by him in conducting his experiments. The more aware he is of the nature of these presuppositions, the more exact is his experimental method" (p. 271). In the same vein, the author holds that "*every statistical hypothesis should be a consequence of a formal theory of nature*" (p. 218). The direction of this proposition is undoubtedly good, but it goes much farther than the reviewer would care to go.

In his approach to control, the author emphasizes the stochastic limit again: " . . . *an experiment is said to be controlled if we state all the formal conditions under which a mathematical function of a series of observations approaches a limit stochastically*" (p. 182). This strong definition is then used in: "*No question of fact can be said to have meaning unless there exists a controlled experiment for its answering*" (p. 183). The reviewer feels that this is a roundabout way to say that no question of fact has a meaning.

In discussing the adequacy of formal probability theories, the author seems to confuse "determination in theory" and " determination in practice." He demands: "Let $O(x_1, x_2. \ldots, x_n)$ be any random sample, with known elementary probability law; let t be any statistic of the sample with degrees of freedom at least 1; then the theory should bee able to state the elementary probability law of t" (p. 19), and then he asserts (pp. 19 and 30) that this demand has not been met in the present probability theory. While the mathematical statistician may not be able to provide a compact and usable answer to many problems of distribution, he can provide a systematic and finite process for determining the distribution within any preassigned limits. In the author's sense of "answer" it seems to the reviewer that modern probability theory provides "answers" to all problems of distribution involving a finite number of observations.

In the author's discussion of the philosophy of science, this reviewer was struck by the statements that (i) "We may find it methodologically profitable to keep contradictory tenets within science" (p. 192); (ii) "There is no true beginning-point to science" (pp. 209–210); (iii) "The time has come to recognize the circularity, or spiral form, of science, and the complete interdependence of the sciences" (p. 216); (iv) "Hence, science demands a science of efficiency, and cannot establish such a theory within psychology or the

science of social groups. The science of ethics, for such we call the measure of loss, must on the one hand belong to experimental science, and yet not be an aspect of any of the special disciplines now recognized" (p. 250). With the first three of these the reviewer is in hearty accord, on the fourth he feels uninformed.

The real difference between the author and the reviewer is in their approach to models, whether mathematical or formal. The author is prepared to take a model on its face value, apparently without consideration of its weak points. It does not seem to the reviewer that this is how science has made its great gains by the use of models. It is by combining a working model with more a detailed, and probably unmanageable, model which indicates the soft spots of the working model that science has progressed.

Passing now from matters of opinion to matters of fact, there are a few specific points. On page 7, the author states that, when the null hypothesis holds and sample size increases to infinity, "t will have a limiting value of zero." This is incorrect. On page 35, the author states that the problem of confidence intervals for means of later samples is "the so-called problem of Tolerance Intervals." This is a slip. On page 257, the author suggests that the best test is obtained by minimizing the integral of the risk over the parameter space. This is not invariant, and of doubtful utility. A similar difficulty occurs at the top of page 211. The wording of the next to the last paragraph on page 9 and the two-valued use of n on page 16 seems sloppy. On page 12, the author asserts that "we could—find a best test" for slippage with an arbitrary continuous distribution. The reviewer would be interested in the definition of "best" and the resulting test.

The book is singularly and pleasantly free from typographical errors—the only ones noted were "procedure" on page 207 and "m_1/m_2" for "m_2/m_1" on page 211—and is excellently printed and bound.

The reviewer would not have taken so much space to review a book he judged of little use. Although he disagrees with the author on almost all the really basic points, he plans to use the book in connection with a course in the design of experiment this fall.

Quality Control: A Manual of Quality Control Procedure Based Upon Scientific Principles and Simplified for Practical Application in Various Types of Manufacturing Plants. *Norbert L. Enrick* (Associate Professor of Management, Southwestern Louisiana Institute, Lafayette, La.). New York 13: Industrial Press (148 Lafayette St.), 1948. Pp. vi, 122. $3.00. (Brighton 1, England: Machinery Publishing Co., Ltd. [148 Lafayette St.].) *Two reviews follow:*

Review by J. H. Curtiss
Chief, National Applied Mathematics Laboratories
National Bureau of Standards, Washington, D. C.

According to the Introduction, this book is intended for practical men in inspection who do not want to be bothered with "higher mathematics," but who would like to have statistical quality control explained in

simple terms. "Higher mathematics" here means anything beyond grade school arithmetic. The spirit of the book is perhaps best conveyed by reporting the fact that the presentation of control charts for averages is so arranged that in setting up the control limits, the mean range of a set of samples never has to be multiplied by any factor more complicated than unity!

Thus the author has imposed rather severe conditions of limited visibility on the flight of his muse. The result is a sort of minimum cook-book of statistical quality control recipes, supplemented by some practical advice on the management aspects, and by some rather sketchy and disjointed remarks on tolerances and gages. An elementary discussion of the underlying theory is also given in a few pages at the end of the book.

The statistical quality control recipes occupy the first 45 pages or so of the book, with a little additional statistical material (mainly on "compressed limit gaging" and statistical study of tolerances) scattered through later chapters. The two main statistical techniques discussed are lot-by-lot inspection, using sampling by attributes, and control charts for averages and ranges. There is no treatment of charts for numbers of defects and for proportion defective.

At the outset of the chapter on lot-by-lot inspection, the author promises to demonstrate later that one should not use lot-by-lot inspection on inspection lots containing less than 300 items, but the reviewer was unable to find the demonstration of this theorem in the ensuing text. Sampling tables are given in the form of double entry tabulation, one argument being lot size range, and the other "allowable per cent defective." There are two tables for discrete items, one of them containing sequential sampling plans, and the other single sampling plans with operating characteristics similar to the sequential ones. These tables are supplemented by two roughly parallel ones for use on continuous products. Although no credit is specifically given in the text, the sequential sampling tables are copied from the "Inspection Handbook on Sampling for Quality Control," QMC-M605–15, published by the Office of the Quartermaster General in 1945. Presumably the other tables are taken from material developed for later editions of the QMC Handbook.

The concept of "allowable per cent defective" used in this book seems to be a sort of mixture of Average Outgoing Quality Limit (AOQL) and Acceptable Quality Level (AQL) as these terms are now used in the technical literature. Mathematically, the "allowable per cent defective" ascribed to each plan is approximately equal to its AOQL, a fact implied by the brief elementary discussion of the theory of the plans given at the end of the book. The instructions and examples, however, seem to handle the "allowable per cent defective" as if it were an AQL.

The control chart for averages is set up as a test of the compound hypothesis that the population mean μ lies in the range $T_1+3.1\sigma \leq \mu \leq T_2-3.1\sigma$, where T_1 and T_2 are preassigned lower and upper tolerances for individuals and σ is the population standard deviation. The test is carried out with the

arithmetic mean of a sample of size 3 to 5, using a level of significance corresponding to a 2 σ tail of the distribution of the mean. That is: the control limits are given by $T_1+(3.1-2/\sqrt{n})\sigma$ and $T_2-(3.1-2/\sqrt{n})\sigma$. If $3 \leq n \leq 5$, the theoretical mean value of the sample range (assuming normality) is roughly equal to $(3.1-2/\sqrt{n})\sigma$, so the very simple practical rule for finding the control limits mentioned in the first paragraph of this review is obtained.

This type of control chart of course differs from the standard Shewhart control chart for averages from the viewpoint of both engineering and mathematical theory. An obvious minor disadvantage (which, however, may be a grave one for the intended users of this manual) is that a pair of tolerance limits must be given before the recipe can be carried out. A major disadvantage is that a fundamental Shewhart control chart doctrine is ignored: a principal goal in quality control is the achievement of a state of statistical control about stable population values of μ and σ. In the present type of chart, μ is theoretically permitted to wander about at will between the limits given above. But the book is written primarily for the line inspector, and not for management, nor for quality control engineers; and perhaps a control chart which places first emphasis on the immediate avoidance of nonconforming product, as this one does, is the right one to present under the circumstances. The control chart for the range is given the orthodox treatment.

In his effort to be brief and clear, the author omitted some points which the reviewer considers rather essential for the proper application and interpretation of even the few simple techniques here treated. No operational meaning is given to the words "random sampling," which are explained in a circular manner by simply repeating the word "random" in a couple of different contexts. Rational subgroups are not adequately discussed in connection with control charts. In the discussion of tolerance ranges, the correct *location* of the range, as determined statistically. is ignored, and only the width of the range is discussed.

As stated before, the technical material on statistical control takes up a little less than half of this 120 page book, and in the opinion of the reviewer, it could have been presented in an unhurried pamphlet of about 25 pages. The density of thought in the chapters on tolerances and gaging, and in other later chapters, seemed rather low, and the reviewer wondered how much of that material would be new or useful to an experienced inspector. The book certainly does not begin to cover adequately the non-statistical aspects of quality control, and indeed the author would probably disavow any intentions in this direction.

On the other hand, the style is simple and clear; the many examples are well-chosen and informative; and all-in-all, this is a very readable little treatise. It must be left to members of the intended audience, rather than to this reviewer, to judge whether the extra pages were well worth the time and effort. The judgment may very well be in the affirmative.

REVIEW BY E. H. MACNIECE
Director of Quality Control, Johnson & Johnson
New Brunswick, New Jersey

THIS book effectively leads the reader into a simplified quality control program. The program is clearly outlined and stated in such a manner that the shopman is familiarized with the use of the method without the complicated terminology and mathematics so frequently found in books on this subject. Perhaps its greatest service will be in the conditioning of non-technical shop personnel for the acceptance of quality control as a means of achieving productivity in terms of acceptable quality with low waste rather than high production with too much of it finding its way to the scrap pile or the salvage department. Mr. Enrick's book is highly recommended as primary reading for men in industry who want to produce acceptable economic quality.

Traffic Performance at Urban Street Intersections. *Bruce D. Greenshields, Donald Schapiro,* and *Elroy L. Ericksen.* Yale Bureau of Highway Traffic, Technical Report No. 1. New Haven, Conn.: Bureau of Highway Traffic, Yale University, 1947. Pp. xv, 152. Gratis.

REVIEW BY HARRY G. ROMIG
Member of Technical Staff, Bell Telephone Laboratories, Inc.
463 West St., New York 14, N. Y.

THIS report presents a practical as well as a statistical analysis of traffic data covering "the intersections of streets at grade in urban areas." Its importance is readily realized since "one-half of all urban traffic accidents and more than three-fourths of all delays experienced in dense urban areas are related to intersections." The traffic engineer will find much new valuable material in this report, and should have it handy as each page presents important details that require careful study.

The manner of presentation is excellent. The table of contents and the complete index at the end are in sufficient detail to make them satisfactory for ready reference. As much of the descriptive matter centers around the figures and tables, the authors provide a fine descriptive list of each with accompanying page references. The book consists of six chapters and six appendixes. Chapter 1 presents the techniques used in collecting the field data in permanent form for analysis. Photographic devices used are described in sufficient detail that others may follow the same procedures in making other similar surveys. Chapter 2 describes "Starting Performance at Signalized Intersections." Practical and theoretical solutions are presented. Chapter 3 covers "Deceleration of Motor Vehicles at Street Intersections." The findings of the study are given in a simple but forceful manner. Chapter 4 presents the "Behavior Patterns at Unsignalized Intersections." The Methods of Analysis are described together with the Detailed Analysis

of Specific Aspects of Behavior. Chapter 5 considers "Highway Traffic and the Theory of Probability." The nature of the distributions found is discussed and it is shown that the Poisson series may be used effectively in the analysis as the Poisson distribution appears to fit the data. There are two distinct parts to this chapter, one describing the General Theory and a second covering the Theory of Random Distribution applied to Signalized Intersections. Chapter 6 describes "Typical Traffic Problems" and indicates their solutions. An excellent summary of each chapter is given at its close in all cases but the First and Fifth. Chapter 1 has no summary, while Chapter 5 has a summary for the general theory and also a summary covering the case for random distributions. Also six valuable appendixes have been provided dealing in mathematical relations, tables and important theories that have been expanded in detail to supplement the main report.

Throughout the study, in taking pictures or making graphs, frame time intervals of 1/88 of a minute were used. This makes it possible to express velocity directly in miles per hour if measurements of distance between time intervals are expressed in feet, i.e., an automobile traveling 5 feet in one such time interval has a velocity of 5 mi./hr. Pictures were taken at sufficiently high elevations to provide a view of the intersections studied and timing devices were included to permit ready identification of the different frames. Later, it was possible to study each frame individually or a run of frames to properly analyze the different conditions under study. In addition to the splendid charts provided in this report, 9 photographs are included showing the intersections involved, and the projector and mounting.

In studying starting performance at signalized intersections, attention was given to three factors: (1) time required for vehicles to commence motion, (2) distances reached by vehicles in given time intervals after starting, and (3) spacing between vehicles. Small trucks are treated the same as passenger cars, but buses and large trucks are studied separately. Where no signal occurs at an intersection or one street has a "Stop" sign, collision points were selected in the middle of the intersections. Velocity, delay factors, reactions of different drivers, and other factors were considered. Time value to the collision point was found to be the main criterion by which drivers decide to take precedence over vehicles approaching on the cross street.

The report added much to its value by including the range of variation indicated by the data for the various situations covered. Its last two chapters discuss the application of probability theory to the problem considered and indicate the use of the Poisson exponential distribution. The authors indicate that in 1934 "the theory that traffic follows a random distribution was assumed by Mr. John P. Kinzer in an article in which he calculated the probability of any car picked at random going a mile without interference or delay on a two lane road with a given volume of traffic." The authors develop the theory and show that with the exception of one second spacings the Poisson theory fits their data fairly well. The exception is due to the desire of drivers to avoid rear end collisions. It is possible to apply the theory and

obtain reasonable solutions to many traffic problems, which formerly defied solution.

Many numerical examples are given and also relations covering the solution of different varieties of traffic problems. Approximate relations are given for use in solving problems when the work of computation becomes too difficult for obtaining the exact theoretical solution. Chapter 5 covers the theoretical treatment and Chapter 6 presents solutions to a number of typical traffic problems. To readers other than traffic engineers it would have been helpful to have presented in an introduction, or in Chapter 1, the typical traffic problems that are to be solved. Even after reading the report several times, it is not clear how the timing of red and green signals are obtained for the most efficient movement of traffic. When should there be a flashing red? a flashing amber? a policeman in control? a Stop signal at only one intersection? a Stop signal at two intersections? Chapter 6 is supposed to provide answers to some of these questions. Those preparing the report were doubtless more interested in the analysis of their results than in delineating how these results can be applied. The last example on signal timing is excellent. More applications of this nature should be included. Other reports can be made more valuable by spending a little more time at the beginning and end in showing how to use the findings.

Mathematics of Sampling. *Walter A. Hendricks* (Principal Agricultural Statistician, Bureau of Agricultural Economics, Washington 25, D. C.). A summary of a course of lectures given during the 1947 Statistical Summer Session at Virginia Polytechnic Institute. Virginia Agricultural Experiment Station, Special Technical Bulletin. Blacksburg 13, Va.: the Station, February 1948. Pp. ii, 45. Gratis.

REVIEW BY T. A. BANCROFT
Research Professor of Statistics and Director of the Statistical Laboratory Alabama Polytechnic Institute, Auburn, Alabama

ALTHOUGH the material for the most part is not new, having been taught in survey sampling courses at various statistical centers, in particular at Iowa State College, it presents in published form an introduction to the mathematics of survey sampling. It should be a welcomed addition to the unfortunately small amount of published material available in this rapidly expanding field of statistics. The booklet should be of value as a reference for workers engaged in survey sampling as well as for teachers and students of its theory and practice.

The mathematical aspects stressed are those that are basic to an understanding of the sampling designs and analyses used in actual sampling surveys conducted at the present time and for the most part by various federal agencies and certain universities with strong statistical sections. Since the booklet is concerned with the mathematics of sampling, no attempt is made to discuss techniques of planning, schedule or questionnaire construction, organization, field operations, etc. A good idea of the type of material dis-

cussed can be obtained from the following list of headings: Classical Error Theory, Random Sampling in Practice, Analysis of Variance and the Estimation of Variance Components, Stratified Sampling, Subsampling, Cluster Sampling, Binomial and Multinomial Sampling Variation, The Problem of Nonresponse, Linear Regression, and The Method of Least Squares. A selected but valuable list of references is given in a section on suggested reading. In the reviewer's opinion the value of the booklet would have been greatly enhanced by the addition of sections on: choice of sampling unit, determination of sample size, confidence intervals, double sampling, and variances of totals based on various methods of estimation.

Although the title of the booklet contains the word "Mathematics," no attempt has been made to give either rigorous detailed mathematical proofs or to introduce useful powerful mathematical concepts or machinery to shorten such proofs. Instead the heuristic approach has been used, the details of proofs in many cases being suggested rather than explicitly stated. General theorems, probability distributions, and formulas have been advanced as true because of their analogy with simpler cases. The manner of presentation is understandable since the booklet is a summary of a few lectures covering a broad field. It seems to the reviewer that a valiant effort has been made, by the use of these methods, to make the methodology of survey sampling reasonable to workers engaged in this field who may have a modest background in the elements of mathematical statistics. If such be the case, it seems to the reviewer that on the whole the author has been successful with one important exception. It is the opinion of the reviewer that the fundamental assumptions and limitations involved in setting up various mathematical models, especially in the case of the analysis of variance and of such proofs as indicated in formula (52) and the simpler case at the bottom of page 11, should be given. It is true that an indication of the fundamental assumption of linearity in the analysis of variance model is given in equation (50), but in the reviewer's opinion a greater understanding would have resulted from beginning with the usual assumptions, i.e., $x_{ij} = u + f_i + e_{ij}$, etc., and with detailed definitions, even though the derivation in the latter case is longer.

No mention seems to have been made of the formulas for the variance of a product and the variance of a quotient. For the sake of comparison, it would seem desirable to give a proof of the usual formula found in the literature for the variance of the mean of a sample from a finite population in addition to the one given in the booklet.

There are several typographical errors. Also the reviewer differs with the author on several points of notation. On page 6 in (28) and again on page 15, r has been used in place of ρ. On page 12, Table 2, σ^2 should be added to $K\sigma_c^2$ for the mean square between classes. On page 9, in (41), it would seem more appropriate to replace s^2 by $(s')^2$ since s^2 is defined by (38). Page 13, (49) and (50) should have k_0 and $\bar{x}_i$ respectively on the left sides. In solving for various variance components, greater clarity should result in replacing σ_c^2 and σ^2 in the equations of estimation, by $\hat{\sigma}_c^2$ and $\hat{\sigma}^2$.

Principles of Medical Statistics, Fourth Edition. *A. Bradford Hill* (Professor of Medical Statistics and Director of the Department, London School of Hygiene and Tropical Medicine, University of London, London, England). London W.C.2: Lancet Ltd. (7 Adams St., Adelphi), 1948. Pp. xi, 252. 10s. 6*d*.

Review by Margaret Martin
Assistant Professor of Preventive Medicine and Public Health
Vanderbilt University, Nashville, Tennessee

News that A. Bradford Hill's excellent book on medical statistics is again available, in the form of an enlarged fourth edition, is indeed welcome. The principal changes from earlier editions are the addition of a new chapter on averages, a section on the normal curve, and the expansion of the chapters on frequency distributions and graphs, chi square, life tables, and standardized death rates.

The clarity of the presentation, the emphasis on the meaning and interpretation of statistical results, and the inclusion of numerous examples illustrating the dangers of careless statistical thinking account for the popularity which this work has enjoyed since its first appearance as a series of articles in *The Lancet* in 1937. Medical students, physicians, and other workers in the medical fields who wish to gain an understanding of the principles of elementary statistics will find it most helpful and stimulating.

On the whole the selection of material to be included in this elementary text of nonmathematical character is excellent. The reviewer feels that it would be desirable to have included a table of probabilities for the normal curve to be used in significance tests, especially since such a table is given for chi-square; that in the discussion of significance tests for proportions, more detailed consideration might have been given to the conditions necessary for reasonably reliable application of normal curve theory and to the correction for continuity; and finally, that follow-up studies in which cases are under observation for fractional parts as well as for whole numbers of years might have received more complete treatment. On the other hand, the calculation of the average length of after-life in a study in which life experience is not complete for all patients (p. 173) does not seem to be particularly useful and might, in fact, lead to misinterpretation.

A few minor errors have been noted. In the calculation of the median of grouped frequency distributions (pp. 49–51), the point below which there are $(N+1)/2$ instead of $N/2$ observations, assuming that the observations are evenly distributed over the interval in which the median falls, is obtained. In the diagram on page 65 the intervals labeled one, two, and three standard deviations, respectively, are actually twice this amount. The appearance of a "minus sign" in line 6 of page 74, when algebraic signs have been ignored in corresponding situations in earlier examples (i.e. in a correction term which is to be squared), might cause some confusion to the reader. In the definition of the weighted mean on page 245, some necessary parentheses have been omitted in the numerical example. In the definition of the chi-square

test on page 246, the word "frequencies" would seem to be more appropriate than the word "values."

An Experimental Introduction to the Theory of Probability. *J. E. Kerrich* (Senior Lecturer in Mathematics, University of the Witwatersrand, Johannesburg, South Africa). Copenhagen, Denmark: Einar Munksgaard, 1946. Pp. 98. Paper.

REVIEW BY J. F. KENNEY
Associate Professor of Mathematics, University of Wisconsin
623 West State St., Milwaukee 3, Wisconsin

UNDOUBTEDLY many teachers have had experiences similar to those of the author in presenting lectures on elementary statistics to mixed classes of students and colleagues who vary widely in mathematical preparation and whose interests lie in diverse branches of science. "It is a most interesting problem," the author remarks, "to design lectures suitable for such a class." An opportunity to design suitable material on one topic came to him when he found himself interned (for his own safety, as a British subject) by the Danish government during the recent war. Thus he had the leisure and patience to conduct the simple but extensive experiments on random events that comprise the subject of this book. The main experiments consisted of spinning a coin 10,000 times and drawing 5,000 times two ping-pong balls out of four of which two bore a red trade-mark and two a green trade-mark. (The drawings were made by a fellow-internee "at a rate of about 400 times an hour with—need it be stated—periods of rest between successive hours.") Also an experiment equivalent to tossing a biased coin was performed with a small wooden disc coated on one face with lead.

Various results from these experiments are recorded in tabular and graphical form. Data are analyzed both in the large and in sub-sequences with respect to various ratios such as m/n where m denote number of heads and n number of spins, and $m_2/(m_1+m_2)$ where m_2 denotes the number of times that green was second in the m_1+m_2 experiments in which red appeared first. The analysis leads to a body of ideas, namely, a mathematical theory which describes the observations. Thus the author arrives at the tools of pure mathematics. Using appropriate symbols he discusses complementary, joint, mutually exclusive, compound, and conditional events, the addition and multiplication principles, and the binomial distribution. The normal distribution is mentioned briefly and an introduction is given to the notions of estimation and confidence intervals.

In the reviewer's opinion the author has admirably achieved his objective as stated in the Foreword: "In this book, a little ground is covered thoroughly and great pains are taken to try to present a clear picture of the physical significance of a mathematical probability. With this background the student will be better equipped to study the many texts which deal with 'pure' theory based on a system of axioms." And his hope is well founded

when he says: "It is hoped that students of these pages will never have to reject any of the ideas given here, no matter how much they may refine them as their knowledge of the subject grows."

Statistical Methods in Medical Research: I, Qualitative Statistics (Enumeration Data). *Donald Mainland* (Professor of Anatomy, Dalhousie University, Halifax, Nova Scotia). Reprinted from the *Canadian Journal of Research, Section E, Medical Sciences* 26 (1): 1–181 February 1948. Ottawa, Canada: National Research Council of Canada, 1948. Pp. 181. Paper. Apply. *Two reviews follow:*

REVIEW BY JOHN W. FERTIG
Professor of Biostatistics, School of Public Health of the Faculty of Medicine Columbia University, New York 32, N. Y.

THIS article is essentially an expansion of Chapters 2 and 3 of the author's *The Treatment of Clinical and Laboratory Data* (Edinburgh, Scotland: Oliver and Boyd Ltd., 1938). A detailed consideration is given to small samples of enumeration data in which the normal curve or chi square solutions are not completely satisfactory. Fifty-four pages of tables are presented giving confidence limits for a two-fold classification of enumeration data and probabilities or significant differences for four-fold contingency tables. There is also a table of chi square and one of four place logarithms of factorials. The text covers 103 pages and is divided into an introductory section of 11 pages, one of examples covering 54 pages, and one of explanatory semi-theoretical notes covering 36 pages. Most of the examples are concerned with the comparison between a sample and a population relative frequency and with the four-fold table, including numerous variations of these problems. The problem of non-dichotomous scales is only briefly considered, as is the problem of combining information from two or more samples.

The suggested treatment of the numerous examples consists largely in aiding the reader utilize the tables contained in the article. Practically no treatment of the rationale of the method is given at the time of the discussion of the example. This is reserved for the section on notes. Each example is, however, followed by a series of helpful comments. While this reviewer recognizes that the investigators for whom this presentation is intended are often not very patient with a discussion of the reason for a certain statistical method. he still feels that the incorporation of the section on notes together with the examples would have produced a much better appreciation of the techniques.

It is sometimes difficult to appreciate the reason for the author's preference for chi square, for example, on page 39: "Some investigators still use the standard deviation or standard error, $\sqrt{Npq}$, instead of chi square, for comparison of the sample. This is not to be recommended." It is not pointed out clearly that the correction for continuity can be used for the normal curve as well as for chi square. The author recommends the summation of

chi square values for combining information from several samples, but this method may at times be unduly conservative.

The author has to some extent achieved his goal of classifying certain types of problems relating to enumeration data, and of telling the investigator how to find his problem and a suitable answer for it. The tables supplied are indeed very comprehensive and useful. However, it seems to this reviewer that the approach is too mechanistic and would be unsatisfactory for many investigators.

Review by A. Bradford Hill

Professor of Medical Statistics and Director of the Department
London School of Hygiene and Tropical Medicine
University of London, London W.C.1, England

Dr. Mainland says that his article was devised "to meet the wishes of those who, in the words of one investigator, would say: 'I have a problem on hand . . . Must I spend a month of free evenings reading a book from end to end several times and mastering all details before deciding how to go about solving the problem? I hope not.'" He might have retorted that workers in the medical sciences would not expect to use, say, bacteriological or pathological techniques without mastering the details and is there any reason why statistical processes should be regarded differently? There is, perhaps, at least an excuse. Few workers unless trained in such subjects as bacteriology or pathology will be so bold as to embark upon them; almost all, whatever their subject matter, will sooner or later be faced with statistical data and have to interpret them. Often too, particularly in clinical medicine, their numbers of observations will be small. It is, therefore, legitimate to argue that it is better to give the worker easy access to tests of significance which he may imperfectly understand rather than to let him rely solely upon that "common-sense" which is, in fact, so uncommon.

The serious danger of this procedure, which Dr. Mainland recognizes, is that the worker may come to regard the mathematical tests as the most important part of the statistical methodology and forget that of much greater importance "are, first, the planning of the experiment or observation so that valid inferences shall be obtainable, and, secondly, the interpretation of the results of the mathematical tests." In the experience of the reviewer the latter is the greater risk, that too frequently today there is a tendency to regard "non-significant" as implying guiltless rather than non-proven, "significant" as proven *and therefore due to* a particular causal factor. Dr. Mainland has certainly endeavoured throughout his article to guard against these very undesirable by-products of his plan of presentation.

This plan is as follows. In an introductory section he discusses, briefly, some general principles and definitions—random sampling, probability, confidence limits, the comparison of samples, levels of significance, etc. He

then passes to what is the crux of the article for the investigator quoted at the beginning of this review, namely the working out of 40 examples of medical problems classified so that the worker can choose data and problems comparable to his own and then easily carry out the demonstrated probability calculation. As the article is confined to qualitative statistics the types of problem are mainly the argument from a sample to its population and the comparison of two or more samples (with subsidiary questions that flow from them). A final section of "notes" discusses the underlying principles and methodology in much greater detail and the article concludes with some extremely useful tables. These comprise binomial confidence limits (with graphs as well) over a wide range of size of sample and of values, and also exact probabilities for small-sample fourfold contingency tables—the probabilities for equal samples up to N equals 20 and the significant differences for unequal samples up to N_1 equals 20 and N_2 equals 19. These latter should clearly be of great help to many workers, as will also a table of the logarithms of factorials of numbers up to 1,000 for the calculation of exact probabilities not tabulated. For samples not covered by the tables precautions and rules regarding the use of chi squared have been derived from more than five hundred comparisons between chi squared and the exact method.

A criticism that might be made is that Dr. Mainland is rather inclined to overstate the case for using the "exact" methods he gives—how often, in fact, would the observer be misled by the cruder methods if he were cautious in borderline cases?—and to place considerably too much confidence in the results given by very small samples. While agreeing with him that "no sample is too small for statistical assessment" one may yet, for instance, with a mere handful of sick persons to compare remember their innate variability and Dr. Mainland's own emphasis on the importance of "the interpretation of the results of the mathematical tests." However that may be, this heavy piece of work should certainly help the medical investigator to apply without tears his tests of significance to small samples—though it is unlikely that he will do so intelligently unless he is prepared to take *some* trouble to understand what it is all about.

Mathematical Theory of Human Relations: An Approach to a Mathematical Biology of Social Phenomenon. *N. Rashevsky* (Associate Professor of Mathematical Biophysics, University of Chicago, Chicago 37, Ill.). Mathematical Biophysics Monograph Series No. 2. Bloomington, Ind.: Principia Press, 1948. Pp. xiv, 202. $4.00.

REVIEW BY FREDERICK MOSTELLER
Associate Professor of Mathematical Statistics
Department of Social Relations, Harvard University
Cambridge 38, Massachusetts

RASHEVSKY'S *Mathematical Theory of Human Relations* has the subtitle "An Approach to a Mathematical Biology of Social Phenomena." It is interesting to notice that Rashevsky feels that such topics as distribution

of city sizes, economic interaction of the social group, variations in the class structure of a group, individualistic (capitalistic) and collectivistic (cooperative) societies, history of nations, theory of war, all come within the scope of mathematical biology. This view fits in well with the present breakdown of borders between sciences.

The writings are largely publications of Rashevsky in *Psychometrika* collected in such a way as to provide continuity to the exposition. The continuity is more one of method than of subject matter. The casual reader will find that the topics dodge around rather rapidly. Indeed, the book is something of a hodge-podge. It contains many early thoughts not very thoroughly worked out but apparently put down quickly as they came to mind by a rather prolific but not very elegant writer.

The principal method employed is that of differential equations, used somewhat in the manner of the applied physicist. One has the feeling that the problems were made to fit the mathematics with which the author has been successful in treating other problems, rather than making the mathematics suitable to the problems. Occasionally the author sidles into integral equations but no serious attempt is made to do anything about them. However, the integral equation approach did look rather promising before it was dropped like a hot potato.

In reviewing such a book about human relations, one has to consider the scarcity of mathematical works on this subject outside the fields of economics and population. Naturally statistical methods are widely used in all social sciences, but these are usually employed for descripitve purposes rather than as mathematical models. Occasionally there are statistical or probability models which can be classified as mathematical models in the sense that they try to explain the way certain processes combine to produce a certain outcome, and these models have the property that many different aspects of the situation can be derived from the original set of assumptions. The work of Zipf has been largely the collection of certain kinds of number anomalies with guesses about the sociological meaning of these anomalies, while Stewart working on the same subject seems to be trying to build up a theory leading to these number anomalies from a set of assumptions. Starting from theory which he has developed for another purpose, Rashevsky tries to investigate the distribution of city sizes but is not very successful. Lewis F. Richardson in *Generalized Foreign Politics* (British Journal of Psychology Monograph Supplements, No. 23, 1939) attempts to study the theory of stability of peace between two or more nations largely by the study of the behavior of linear differential equations. Richardson is not quite so ambitious as Rashevsky. He studies conditions which will lead to war and does not attempt to say when war will occur, nor when one side or the other will be defeated, nor how the action will be carried out, while Rashevsky does make attempts of such a nature. The contrast between the work of Richardson and that of Rashevsky is worth noting because the one man takes a single topic and works it very extensively, while the other prefers to handle many topics thinly.

Rashevsky places a critic of his work in a very difficult position. He states, as we would expect any man building mathematical models to do, that none of the models he presents are to be taken as sacred or complete or more than a gross oversimplification of the techniques he has in mind. Further, even when he goes out of his way to get some data and compare his theory with some facts, he claims that no one should take the results of the comparison seriously as supporting the particular theory he has in mind. In every case as far as the reviewer can tell, he regards his examples as "only an illustration" of what a man constructing mathematical models might hope to achieve and improve on if he were to make a careful extensive study of the problem. This attitude makes it difficult for us to know whom Rashevsky wrote the book for. Presumably a man familiar with mathematical models would know something about the kinds of things that might be achieved through the use of mathematics and therefore would not need all these illustrations. The social scientist who does not know himself how to handle mathematical models will probably feel that instead of producing all these illustrations of what might be accomplished, Rashevsky might have done better to take *one* problem and work on it. His attitude might be that one good investigation would win him over. Probably Rashevsky protests more than he means and really feels he has a fairly general approach to many social science problems, and he may even feel that he has produced a good framework for building. In addition, Rashevsky may also feel that the reason so little work has been done by applied mathematicians on social science problems (outside the afore-mentioned fields of economics and population) is that the mathematicians see no way to attack these problems; that if encouragement of the kind he is offering is given, research people may see their way clear to relieving the scarcity of work in this field. It is very possible that this book may have the effect of goading researchers to work in this field, because some may feel that Rashevsky has stated his problems poorly and that too many problems are left wide open by Rashevsky's approach. If the book produced only this effect, the author will have made an important contribution to the development of social science.

The book opens with a "Preface and Explanatory Remarks"—section in which the author gives some arguments why mathematics should be allowed to be used in the study of social phenomena and includes a fairly lengthy criticism of this work by the author. Indeed, anyone wishing to criticize this book will be helped by reading Rashevsky's own criticisms. Chapter 1 considers the nature and effect of the influence of one individual on another and provides a definition of social class. Generalizations are achieved by introducing the notions of distribution of individuals in space, and in time, and the notion of social mobility. It is unfortunate that this important first chapter is not written a little more carefully; for example, on page 3 line 6 the reader is confused between an activity and the intensity of an activity. On page 4 the author has not been careful about his use of absolute value signs or else he has changed his assumptions without informing the reader.

It is likely that the reader will be a little startled to see a multiple integral with definite limits suddenly lose two of its three integration signs with no caution from the author that the single integration sign is to stand for all three, even though it has definite limits attached which differ from those on the other two.

Chapters 1 and 2 are largely confined to a discussion of functions of several variables with no definite form assigned. Averages or expected values are largely used. In the study of individuals grouping themselves into classes the author considers the case of one variable F and defines individuals as belonging to the same class when $(F'-F)^2-\Delta^2<0$, where the F's are the values of the characteristic determining the class structure for the two individuals and Δ is some constant. For large groups of individuals, the extent of the upper class is found by averaging the left-hand side of the above inequality over a portion of the joint distribution of two individuals independently drawn from the distribution of the characteristic. If we know N (F), the distribution of the characteristic, and the number of classes in the society, we can in principle calculate Δ.

In Chapter 3 Rashevsky gives an approximate treatment of the interaction of social classes in which he uses an all-or-none principle, that is, either individuals are "active" or "passive." The active individuals belong to two groups each with a single activity in which they try to persuade the passive members to join them. The problem seems to be to see what kinds of conditions will lead to all the passive individuals occupying themselves with one activity or the other. Rashevsky feels that his results agree in general with the rapid spread of mass hysterias and revolts, and the reviewer feels that the ideas may approximate the results observed in fads, fashions, rumours, or propaganda. It is interesting to note that the differential equations produced are of the same form as Richardson's armament equations. This is not surprising because Rashevsky is dealing with warfare between two groups for possession of a third. Rashevsky's equations simplify more than Richardson's because of an additional restriction. Richardson is not mentioned.

Chapter 4 is an extremely useful chapter from the point of view of the social scientist not well acquainted with mathematical models and the adequacy of various kinds of approximations met with in mathematical physics. In this chapter entitled "A More Exact Treatment of the Previous Case," Rashevsky shows that allowing individuals to distribute themselves on an active-passive scale instead of forcing each individual to assume one or the other end of the continuum can lead to essentially the same results as those given by the more approximate methods of Chapter 3. This procedure can teach the social scientist that a process of simplification in mathematical models does not necessarily lead to the loss of essentials. In other words, one should not use too glibly the pat phrase: "Of course, this treatment is much too over-simplified to be of any real use."

Chapters 5 and 6 deal with economic problems raised by the existence

of persons in a society who are so good at organizing that the whole group can gain if the workers will work under the direction of the organizers, and the organizers are willing to organize the workers.

Chapter 7 might interest the practical man, although the illustrations may seem to him to be tours de force. It suggests how previously developed theory can be applied to estimating the ratio of the population of the capital of a country to the urban population less that of the capital, using the proportion of national income taken in taxes. The population of the capital is taken as an index of the size of the governing class and the urban population as an index of the total number of active individuals in a society composed of three classes: the governing, the organizing, and the passive. The results for Germany, France, and the United States look very good, but for England the capital is too large. The reviewer does not think population of a capital a good enough index of size of governing class to make the example convincing. Another example is the prediction of the incidence of crime from taxes and population density, while still a third example deals with the divorce rate. The fit of the calculated quantities to the observed data is rather encouraging. It should be mentioned that in Chapter 6 by a sequence of crude approximations a formula is obtained for estimating the per capita income of a country in terms of its urban population percentage and its population density. The results do not seem to fit very well in this case. In all these cases Rashevsky feels that the real interest attaches to the fact that certain relations are suggested even by an inadequate theory which then helps us notice such relations when they occur.

Chapter 9 is concerned with two notions of individual freedom: the first concerns economic freedom and is rather suggestive, the second deals with freedom of an individual to choose among many activities and seems to the reviewer to fall flat on its face.

Chapter 10 deals with the distribution of the per cent urban population in a growing society and considers two or three possible assumptions. Data are given showing population of a country against per cent urban. The curve derived for the United States fits the data pretty well, although a straight line would fit them better, but that for Germany is extremely convincing. The data for Russia (7 points) are fitted by a two-branch curve with the aid of arguments about the reform history of the country. The data for Sweden are the most interesting available, but are dismissed with the statement that the theory is inadequate to explain them although the rapid reader might think that the excellently fitting curve shown is a derived one. The reviewer cannot tell whether this curve is derived or not but suspects that it is a free-hand fit.

Chapters 8 and 9 deal with city sizes and do not seem to reach a very successful conclusion.

Chapters 14 and 15 deal with social classes, social mobility, production, and the effects of restrictions. Chapter 17 concerns some consequences of previous theory and the theory is extended to estimate the percentage of

per capita income spent for military purposes in various countries and also the number of inventions by various countries. The theory also suggests that the "influence" of a country is proportional to N^3/S, where N is the population and S the area. Graphs for various countries from 1600 to the present do not violate a reader's intuition about which countries had great influence during this period.

Chapters 19 and 20 are concerned with individualistic as opposed to collectivistic behavior and here Rashevsky draws heavily on G. E. Evans' *Mathematical Introduction to Economics*. The principal result is that individuals may profit more individually by trying to maximize the group satisfaction rather than their individual satisfaction.

Chapter 21, "Some Considerations of the History of a Few Nations," discusses largely in hand-waving terms what happens under various degrees of interclass mobility. In other words, the happenings in several different countries, Russia, China, England, and the United States are talked about in terms of some of the theory, although not really derived from the theory. As in many history books, the discussion of the United States concludes "To what extent this shift toward governmental control will continue cannot be predicted on the basis of the present theory" (p. 180).

Chapters 22 and 23 have to do with physical conflict between groups or nations. The theory developed is one of the variations of Lanchester's Law, although Lanchester is not mentioned.

A few misprints, mostly minor, were noted. The more important are: page 17, equation (9), C should be (; page 18, line 5, $>$ should be $\gg$; page 78, equation (6), delete second equal sign; page 84, equation (16), bar of radical should not cover second term.

The most important thing is that a book has appeared which tries to treat a variety of social problems by means of mathematical models. That the attempts have met with varying degrees of success is not too important. The results given are certainly successful enough to encourage others to make further attempts. Indeed, some of the basic material presented here is worth extending along the lines indicated by the author and worth supplementing with practical numerical examples drawn from data.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44 JUNE 1949 NUMBER 246

ARTICLES

PROCEEDINGS OF THE 108TH ANNUAL MEETING

BOOK REVIEWS

A cumulative Index to Volumes 1–34, 1888–1939, and Annual Indexes thereafter, may be obtained from the Office of the Secretary of the AMERICAN STATISTICAL ASSOCIATION.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 246 JUNE 1949 *Volume 44*

THE CURRENT STATUS OF STATE AND LOCAL POPULATION ESTIMATES IN THE CENSUS BUREAU*

HENRY S. SHRYOCK, JR.,
AND
NORMAN LAWRENCE
Bureau of the Census

EARLY IN THIS CENTURY, the Bureau of the Census was already experimenting with postcensal State and local population estimates based on such symptomatic data as the number of public utility consumers, voting registrations, counts from city directories, and school census figures. Estimates were apparently made by the simple ratio method, and it was decided that they were not so satisfactory as those made by a simple mathematical method. For many years, virtually all the published estimates were linear projections of the last two census figures. Later, when the growth of birth and death registration areas made possible fairly accurate estimates of the national total population, an apportionment formula was used to prorate the postcensal national growth among the States and cities in accordance with their previous relative growth.

Although many estimates by these two methods were published, the Bureau recognized that the average errors were large enough to cast doubt on their usefulness. Sudden changes in population trends during the early years of the depression of the 'thirties made it obvious that these mathematical assumptions could no longer be used. In 1936 the

* Adapted from a paper read by the first-named author before a Customer Administrative School for Public Health Executives and Vital Statistics Registrars at the International Business Machine Corporation, Endicott, N. Y., July 2, 1947.

The authors wish to express their appreciation to Mr. Benjamin Greenberg, who assisted in the preparation of the population estimates given here, and to Dr. Joseph F. Daly, who directed their attention to the new tests of statistical significance in an unpublished thesis by John Edward Walsh of Princeton University.

Bureau published its first series of State estimates by what has been called the "migration and natural increase method." Here the components of population change, natural increase and net migration, were estimated separately, the migration component being derived from a comparison of actual with expected school enrollment. Current State estimates were discontinued after 1937 pending the results of the 1940 census. It had been planned to resume publication after 1940 using an improved migration and natural increase method, based on school data or perhaps some other symptomatic series; but the war changed the situation in several ways.

The program of current population estimates in the Census Bureau received a windfall in the registrations for war ration books in 1942 and 1943. Primarily from this source, several series of State and county estimates were constructed for these years.[1] These were the most comprehensive such postcensal estimates based on empirical data ever published. On the average, they were undoubtedly the most accurate also.[2] With a few exceptions, no population estimates for cities or counties have been published by the Bureau since those for 1943.

The valuable war ration book figures were not, however, an unmixed blessing to producers and consumers of population estimates. Their existence interrupted the experimentation with the above-mentioned symptomatic series, such as school data, that are available in peacetime as well as in war years. A further delaying influence was the prospect of a national sample survey of population in October, 1946, from which population totals for States and large cities would have been forthcoming. The Congress appropriated money for the preparatory work on this survey, but in the following fiscal year it failed to approve the necessary remaining funds. From individual sample surveys in the fall of 1946 and April, 1947, the Census Bureau obtained estimates of total population for selected large cities and metropolitan districts, however. These will be discussed later.

Meanwhile, the demand for State and local estimates was mounting. People were conscious of the sweeping changes that had occurred since

---

[1] U. S. Bureau of the Census. *Population.* "Estimates of the civilian population by counties: May 1, 1942." Series P-3, No. 33, February 25, 1943.

U. S. Bureau of the Census. *Population.* "Estimates of the civilian population by counties: March 1, 1943." Series P-3, No. 38. October 31, 1943.

U. S. Bureau of the Census. *Population.* "Estimated civilian population of metropolitan counties, by single counties: March 1, 1943, and May 1, 1942." Series P-3, No. 40. January 7, 1944.

U. S. Bureau of the Census. *Population—Special Reports.* "Estimated civilian population of the United States, by counties: November 1, 1943." Series P-44, No. 3. February 15, 1944.

[2] Hauser, Philip M., and Tepping, Benjamin J., "Evaluation of census wartime population estimates and of predictions of postwar population prospects for metropolitan areas." American Sociological Review 9 (5): 473–480. Oct., 1944.

the outbreak of the war, and the importance of measuring them. Vital statisticians, market analysts, directors of planning boards, and a host of others had mailed or telephoned their urgent requests. The time of our small estimating staff was taken up with a wide variety of national estimates, and State estimates were next on the list. Within the foreseeable future, we knew we could not make estimates for all the counties and middle-sized cities. Many State health departments just had to have such estimates, however; and the question of giving them technical advice was raised.

METHODS USED RECENTLY

In 1946, Dr. Hope T. Eldridge, then of the Population Division of the Census Bureau, spoke before the Council on Vital Records and Vital Statistics. She described a method for making postcensal county estimates on the basis of school data, vital statistics, and figures on the armed forces. Later she expanded this paper, and it was published as "Suggested procedures for estimating the current population of counties."[3] A second and more detailed method was also outlined in this article.

The second method, slightly modified, was used in making most of the published State estimates for 1946 and 1947.[4] Estimates by both methods were also computed for all the cities surveyed in the autumn of 1946 and the spring of 1947 in order to have independent checks on the inflated sample figures. The prime purpose of the present report is to present what has been learned about these methods and their pitfalls. The Census Bureau would also be greatly interested in hearing of the experience of other workers with these methods.

The methods described in Dr. Eldridge's article may be summarized briefly here. Detailed examples of each method are available on request to the Bureau of the Census. The simpler method, Method I, assumes that the difference between the percentage change in elementary school enrollment for the local area and the national percentage change in population of elementary school age is equal to the percentage change through net migration to or from the local area. The base date for a State, city, or county may be November, 1943, or any other date for

[3] Bureau of the Census *Population—Special Reports*, Series P-47, No. 4, April 30, 1947.

[4] U S Bureau of the Census. *Current Population Reports. Population Estimates.* "Estimates of the population of the United States, by regions, divisions, and States: July 1, 1940 to 1947." Series P-25, No. 12. August 9, 1948. This release also contained revised estimates for July 1, 1940 to 1945. The earlier State estimates for July, 1944 and 1945, were to a considerable extent simply extrapolations of those for November, 1943. Subsequent data on natural increase and shifts between civilian and military populations had been utilized, but net interstate migration had been estimated from that for the period 1940 to 1943.

which the total population is available. Other figures needed to complete the estimates, namely, natural increase, persons away in the armed forces on the base and estimate dates, and members of the armed forces stationed in the county on the estimate date, are handled in a fairly direct manner.

Method II differs from Method I only in a more elaborate technique of estimating net migration from the school data. By this technique, net migration is measured on the basis of the difference between the population of elementary school age as estimated from the school data and the expected population of that age had there been no migration since the base date. The expected population is computed by applying survival rates from an appropriate life table to the population cohort at the base date that became the population of elementary school age at the estimate date. This "aging" process must also be applied to cohorts of births occurring after the base date when estimates are being made for the later postcensal years. To illustrate, the expected population 6 to 13 years old on April 1, 1947, assuming no migration, comprises survivors of births between April 1, 1940, and March 31, 1941, as well as survivors of children under exact age 7 on April 1, 1940. Since data on the population by age are required for the base date, it is usually necessary to go back beyond 1943 to the last census. The "actual" population of elementary school age with which the expected population is compared is computed from the school datum for the estimate date by means of an appropriate ratio based on past experience. Some assumption must be made about the relation of the rate of net migration of the school-age population and that of the population of all ages, the simplest assumption being that the rates are the same. This is a critical assumption and is discussed more fully below.

COMPARISON OF METHODS

Comparisons have been made between estimates made by these methods and population figures from other sources. These sources include the following: (1) Special censuses (complete counts) taken by the Census Bureau at the request and expense of local governments; (2) Censuses of Congested Production Areas taken in 1944 by the Bureau in selected metropolitan counties—these were sample censuses with a coefficient of variation of less than one per cent for the total population of each Congested Area; (3) State censuses; (4) Estimates made from the registration for war ration books.

Estimates made by Methods I and II have been compared with population standards of these sorts for 48 States, 25 cities, and 102

counties. Some cities and counties appear more than once in the list, for different dates. Some of the standards are complete counts. Other standards are subject to a fairly large amount of error; but, on the whole, they should be quite accurate. The population standards for States are the November, 1943, estimates based on War Ration Book Number 4 and published in Series P-44, No. 3. For the 25 cities, the standards were as follows: Ration Book estimate, 7; State census, 7; Census of Congested Production Areas in 1944, 9; and Census Bureau special census, 2. The standards for the 102 counties were distributed as follows: Ration Book figure, 43; State census, 42 (4 in Florida, 35 in Kansas, and 3 in Massachusetts); Census of Congested Production Areas, 16; and Census Bureau special censuses, 1. Similar population standards were available for many other cities and counties, but school data had not been collected for them. All figures are for the civilian population. The estimate dates range from 1943 to 1947.

The results of these two empirical methods may also be compared with those from two methods that do not use local postcensal data. The first of these latter is simple arithmetic projection. The total population of each area was projected on the basis of the 1930 to 1940 trend. This figure was converted into one for the civilian population by assuming that the area had the same proportion of its *de jure* population in the armed forces as the national population did at the same date. The second method is the apportionment method used by the Bureau of the Census in the early 'thirties. It assumes that the area had the same share of the national increase in the postcensal period as in the last intercensal period. Because the civilian population decreased over most of the period in which we are interested, the apportionment method could not be applied in the form just described. We modified the usual method by using the total population, which as a final step was converted into the area's civilian population, again by assuming the same proportionate loss to the armed forces by each area as was lost by the United States as a whole. (In the case of areas that had lost population between 1930 and 1940, it was assumed that there was no further decline in the *total* population after 1940. This total was reduced as just described to get an estimate of the postcensal civilian population.)

There is not space to present here the various estimates and their deviations from the population standard for each of the 175 States, cities, and counties. A summary of the deviations is given in Table 1, however, for each type of area. Happily, the extra work of the school-data methods of estimation appears to be decidedly worthwhile. In

TABLE 1

SUMMARY OF PERCENTAGE DEVIATIONS FROM POPULATION STANDARD OF ESTIMATES BY FOUR METHODS FOR SELECTED AREAS AND DATES
(See text for description of methods)

Item	Number of Areas	Method I	Method II	Arithmetic Projection	Apportionment
Average percentage deviation (disregarding sign)					
All areas	175	7.06	6.48	12.08	12.17
States	48	4.50	3.08	6.14	6.09
Cities	25	6.76	5.13	12.89	11.34
Counties	102	8.34	8.40	14.67	15.24
Root-mean-square percentage deviation					
All areas	175	9.43	9.29	16.89	16.75
States	48	5.32	4.29	7.44	7.29
Cities	25	7.75	7.35	14.54	12.96
Counties	102	11.16	11.23	20.31	20.37
Deviations of 10 per cent or above					
All areas	175	43	40	82	83
States	48	3	2	10	9
Cities	25	7	3	16	13
Counties	102	33	35	56	61
Positive deviations					
All areas	175	92	92	70	75
States	48	32	29	28	28
Cities	25	9	10	1	2
Counties	102	51	53	41	45

each type of area, the arithmetic mean of the percentage deviations disregarding sign is lower for both Method I and Method II than for either of the two mathematical methods, arithmetic projection and apportionment. (This fact does not gainsay that for some individual areas the percentage deviation was lower on the part of the mathematical than of the school-data method.) Average "errors" of 6 or 7 per cent over a period of 3 to 7 years by the better methods may not seem satisfactory for some purposes, but at least they compare favorably with the 12 per cent that would have resulted from the use of conventional methods.

It is of interest to investigate whether any of the differences between the average per cent deviations shown in Table 1 might reasonably be

attributed to chance variation. In computing the significance of the average percentage deviation for one method as compared with another, one should take account of the correlation between the deviations by the one method with those by the other. It is not surprising that there is a correlation between the deviations by Methods I and II and also between those by arithmetic projection and apportionment. (For all areas, there is a correlation coefficient of +.82 between the first two methods and of +.95 between the last two.) It is perhaps surprising, however, that there are also significant positive correlations between the deviations of a school-data method and those of a mathematical method, r ranging from +.36 to +.47. All correlations are positive within each type of area, States, cities, and counties. In other words, since the mathematical and empirical methods have no common assumption, a population that is difficult to estimate by one type of method also tends to be difficult by the other.

In testing the significance of the difference between the averages of two sets of deviations, we have in all cases allowed for the positive correlation by getting the difference between paired percentage deviations (disregarding sign) from two methods and testing whether the mean difference is significantly larger than zero. We thus obtain for all areas combined:

	t	P
Between Method I and Method II (7.06 vs 6.48)	1.98	.05
and Arithmetic projection (7 06 vs 12.08)	5.91	<.001
and Apportionment (7.06 vs 12.17)	6.24	<.001
Method II and Arithmetic projection (6.48 vs 12.08)	7.02	<.001
and Apportionment (6.48 vs 12.17)	7.15	<.001
Arithmetic projection and Apportionment (12.08 vs 12.17)	0.34	.7

Thus the superiority of either of the school-data methods over either arithmetic projection or apportionment is almost certainly not due to chance. The greater accuracy of Method II than of Method I may be deemed of border-line statistical significance as far as this test is concerned.

There may be some doubt about the propriety of combining all three types of areas in the above examination. Table 1 shows, however, that the populations of States, cities, and counties are alike in being more accurately estimated by a school-data method than by a mathematical method. For States, cities, and counties considered separately, the probability that these differences arose by chance alone is always less than .01 and usually less than .001. Method II is significantly more

accurate on the average than Method I only for States. Its lower average "error" for cities could have occurred by chance about 1 in 10 times, and for counties the average accuracy is about the same.

Two measures presented in Table 1 stress the larger percentage deviations from the population standard, whether positive or negative. The square root of the mean-squared percentage deviation when used as a criterion does not change our evaluation of the relative accuracy of the four methods very much. Percentage deviations of 10 per cent or more are next shown, and once again the school-data methods with their separate handling of migration and natural increase appear more accurate than the mathematical methods. By neither criterion, however, is Method II significantly more accurate than Method I.

The tests employed above, of course, depend for their validity on the assumption that we are dealing with functions of normally distributed variables. The failure of these tests to differentiate conclusively between the merits of Methods I and II might, therefore, be due to a failure of the normality assumption. To check on this point a test must be used in which the level of significance does not depend on the distribution law of the variables concerned. Recent developments in the theory of order statistics have provided such tests.[5] The application of such a test indicates that estimates prepared by Method II have an average absolute deviation that is definitely smaller than that for estimates prepared by Method I. When the estimates for States, cities, and counties combined are considered the difference proved significant at the .02 level. For States alone the difference proved significant at the .005 level, and for cities alone at the .05 level. For counties alone the difference was not significant. Accordingly, it appears that the additional work required in using Method II instead of Method I will often be justified.

For the cities and counties, the deviations of the estimates from the school-data methods are more evenly divided between positive and negative values than are those from the other methods. This situation is not true for the States. The higher rate of growth in the 'forties than in the 'thirties is particularly apparent in the nearly universally negative deviations for cities shown by the mathematical methods.

In the estimates made by Method II, the school datum was translated into an estimate of the actual population of elementary school age by applying the 1940 ratio of the two statistics. Since this ratio had been variable prior to 1940, it was supposed that the accuracy of

---

[5] John E. Walsh, "Some significance tests for the median which are valid under very general conditions." Unpublished doctoral dissertation, Princeton University Library, Princeton, N. J.

the final population estimates would be improved if the value of this ratio were extrapolated in some other way to the given postcensal date. We did this in two ways: (1) linear extrapolation of the trend in the ratio between 1930 and 1940; (2) logarithmic extrapolation. The latter assumes that the amount of change in the ratio has been progressively smaller. These two methods of extrapolation have been named Method IIa and Method IIb.

In Table 2, the final results (in terms of percentage deviation of the estimate from the population standard) from Methods IIa and IIb are compared with those from the original Method II, in which the ratio under discussion was held constant. A consistent series of school data was not available from 1930 to the estimate date for all areas, but the methods could be compared for a total of 132 States, cities, and counties.

TABLE 2

SUMMARY OF PERCENTAGE DEVIATIONS FROM POPULATION STANDARD OF ESTIMATES BY THREE VARIATIONS OF THE "MIGRATION AND NATURAL INCREASE" METHOD FOR SELECTED AREAS AND DATES

Item	Number of Areas	Method II	Method IIa	Method IIb
Average percentage deviation (disregarding sign)				
All areas	132	6.20	7.13	7.68
States	32	3.38	4.67	4.72
Cities	19	3.98	4.59	4.53
Counties	81	7.83	8.70	9.58

It may be readily observed that these refinements of Method II do not improve on the original. It would appear that there was either very little change in the ratio of elementary school enrollment to population of elementary school age after 1940 or else the change was of a quite different nature from that between 1930 and 1940. Until at least 1950, we may safely spare ourselves the extra work involved in these particular refinements of Method II. On the other hand, the extra work of Method II as compared with Method I seems to offer promise of more accurate population estimates.

A second source of data that may be used to assess the comparative accuracy of Methods I and II is available in the results of sample surveys taken by the Bureau of the Census in October and November, 1946, and in April, 1947. In 19 cities twice the coefficient of variation for the sample estimate of the civilian population of the city concerned

was about 5 per cent. Forty-two other surveys were based on smaller samples and are subject to still larger sampling errors. Moreover, the sample estimate for each of the 61 surveys was subjected to an upward adjustment of about 4 per cent on the assumption that this adjustment would make the results more comparable with those that would be obtained by a complete census. Evidence from national sample surveys indicates that an upward adjustment of about 4 per cent is needed to make the national Current Population Survey estimates comparable with those which would be obtained from a complete census. The case for applying the same adjustment to the results of local surveys, although convincing, was not altogether conclusive; and, in any event, the adjustment may well have been too high or too low for particular cities.[6] Consequently these sample estimates cannot be regarded as having the same status as the population standards referred to previously. We can merely examine the general consistency of the sample-based estimates with the result for each of the two synthetic methods. Such comparisons are presented in Table 3.

TABLE 3

SUMMARY OF PERCENTAGE DEVIATIONS FROM SAMPLE CENSUS FIGURES OF ESTIMATES BY METHODS I AND II: 61 SELECTED CITIES, 1946 AND 1947

Item	Method I	Method II
Average percentage deviation (disregarding sign)	7.6	4.9
Root-mean-square percentage deviation	9.3	6.3
Deviations of 10 per cent or above	15	6
Positive deviations	19	32
Negative deviations	42	29

Measured by both the average deviation and the root-mean-square percentage deviation, the more elaborate Method II gave results more consistent with the sample estimates than did Method I. The respective average deviations were 4.9 and 7.6 per cent; the respective root-mean-square percentage deviations were 6.3 and 9.3 per cent. These comparisons are based on 61 cases. When we pair the deviations (disregarding sign) from the sample estimate for Method I and Method II, compute the mean difference of the paired deviations, and examine by the t-test whether this mean difference (2.7 per cent) is significantly different from zero, we find that the probability that the deviations from the two methods could have come from the same universe is less than .001.

[6] If this adjustment had been as little as 1 per cent, the conclusions set forth below would not have been changed materially.

A very low probability was also obtained by the previously cited Walsh test.

Our preferred school-data estimates also come much closer than the simpler school-data estimates to an equal distribution of positive and negative deviations from the sample estimates. Method II has 32 positive and 29 negative deviations, whereas Method I has 19 and 42.

Thus the evidence seems to confirm our choice of Method II. Of course, some third method might be even better. Where the school-data estimate and the sample estimate for a city agree, we have some feeling of confidence in both since they are from entirely independent methods. Where the estimates disagree, either or both may be in error.

EVALUATION OF SCHOOL DATA

Inasmuch as both Methods I and II depend largely on school statistics for the measurement of net migration, serious errors of estimation can creep in at the very beginning of the estimating process unless the school data used are assembled with the main purpose always in mind. In either method school statistics are needed to approximate as closely as possible the number of children in a fixed age group. Some types of school data do so better than others. For example, statistics relating only to children of compulsory school age are more useful than those relating to all school children regardless of age since many of the latter go to school only by choice. Thus, data on enrollment by age are to be preferred over data on enrollment by grade for the purpose of making population estimates. If the former are not available, as is usually the case, grade enrollment figures are next to be preferred over other types of pupil statistics. But when such figures are used they should be restricted to the elementary grades, which comprise pupils most of whom are within the ages of compulsory school enrollment.

Two other types of school statistics are usually available. The first, average daily attendance, should be avoided on principle. It is obvious that an increase in average daily attendance may reflect no increase in enrollment or population but merely improved school administration or better weather during the school year. Over the long run, improved economic conditions or changes in parental attitudes conducive to more regular attendance at school may also be important factors. As to the second of these types of data, the school census, experience has taught us to proceed cautiously before using it to estimate population. It is no simple problem to enumerate the population of an area with reasonable

completeness even when trained enumerators do the job under well organized administrative controls. Yet, in most instances, school censuses are taken by school teachers, policemen, or unpaid volunteers, often without the geographic and other administrative controls needed to assure reasonably complete enumeration. As a result, school census data frequently vary in accuracy from year to year making it difficult to estimate the population in an age group by reference to the relationship between the school census and the United States census. Children outside the compulsory enrollment ages are especially likely to be underenumerated relative to a federal census. Frequently school census data are not tabulated by age so that the totals shown comprehend several age groups among which the completeness of enumeration varies greatly. Instances have even come to our attention of reports involving the publication of mere estimates improperly labeled as school census returns. In view of these difficulties, the school census has only rarely been used by the Census Bureau, and then only after careful investigation of the data.

A series that closely reflects the size of a fixed age group may nevertheless contain a defect from the standpoint of measuring net migration for a given area. One such situation occurs when the area for which the estimate is being prepared is not the same as the area for which the school statistics are compiled. School districts, it is true, are legally constituted areas, but they are not always the political areas for which population estimates are desired. Thus school statistics for the San Antonio Independent School District include some pupils who do not live in the city of San Antonio and exclude some who do live there. To use the local figures, a school-by-school tabulation was required to eliminate those not in the city. Another instance in which the area of jurisdiction of the school system is a complicating factor is found when a city has annexed one or more suburban areas since the last decennial census. The legal status of the annexed area changes at an instant in time, but the necessary administrative arrangements to consolidate the annexed school into the city school system may take some time to complete. It may be a year or more before the pupils in the annexed area are reflected in the statistics for the city.

Another stumbling block in the use of school statistics is a concealed inconsistency. Frequently a series of school data appears reasonable and consistent, but subsequent investigation reveals an administrative act which renders the figures not only useless but misleading. Thus, in one instance, a sudden increase in elementary enrollment proved to represent nothing more than a change from a seven- to an eight-grade

elementary school system. There are several different ways in which such a change-over might be made; each will leave its impression on the statistics. In the instance mentioned, the impact of the change-over was felt all at once, creating a suspicion on our part that could be checked. If the impact is spread over many years, it may not be noticed and the data may appear to indicate a gradual population increase when all that is involved is an administrative change in the organization of the school system. Similarly the opening of kindergartens or the unannounced introduction of kindergarten pupils into the enrollment statistics will produce a misleading increase in a school data series. (In general, kindergarten pupils should be excluded from the school statistics because attendance for them is not compulsory. For similar reasons, children less than 7 years old, or at least those less than 6, should be omitted when the data are classified by age.) Still another illustration of essentially the same situation is the omission from the school data series of pupils in special or ungraded classes. The omission might become a matter of consequence if the school system is in the process of expanding such programs and is steadily reclassifying its pupils and assigning them into these classes from the regular classes.

One factor about which statisticians have been apprehensive has usually proved to be unimportant. This is the presence of nonresidents in the local schools. Nonresidents, however, are seldom found in elementary classes because of the travel hazards for young children. Children from out of town are more frequently found in high schools; this fact is another reason for not using high school enrollment data.

The extent to which such circumstances as those mentioned can affect the population estimates may sometimes destroy whatever value inheres in the method of estimation. One of the concealed inconsistencies cited above resulted in an estimate of population 14 per cent greater than that obtained when the inconsistency was eliminated. Anyone using school statistics to estimate population must not only be aware of the requirements for data imposed by the estimating technique but must also be a statistical detective, going behind the data to the operating system from which they were derived, and understanding the changes taking place in educational administration as they affect school statistics.

In many areas, a substantial proportion of the population of elementary school age does not attend public school but receives its education at parochial or other private schools.

The importance of using enrollment figures that include private as well as public schools may be demonstrated by comparing estimates

based on total enrollment figures with those based on public enrollment. From the 61 cities of Table 3, 47 in which private enrollment constituted at least 10 per cent of total enrollment were selected. Two sets of estimates were then made by Method II. The results were compared with the population returned by the sample censuses already discussed.

TABLE 4

SUMMARY OF PERCENTAGE DEVIATIONS FROM SAMPLE CENSUS FIGURE OF ESTIMATES BY METHOD II, USING TOTAL ENROLLMENT AND PUBLIC ENROLLMENT: 47 SELECTED CITIES, 1946 AND 1947

Item	Using Total Enrollment	Using Public Enrollment
Average percentage deviation (disregarding sign)	4.7	8.9
Root-mean-square percentage deviation	5.8	10.1
Deviations 10 per cent or above	3	19
Positive deviations	23	8
Negative deviations	24	39

When parochial, and sometimes other private, enrollment is taken into account, the population estimates are significantly more consistent with estimates obtained by sample censuses. The average absolute percentage deviation for the estimates based on both public and private enrollment (4.7 per cent) was significantly smaller ($P<.001$) than the average based on public enrollment alone (8.9 per cent) by either the t-test or the test devised by Walsh. The improvement is so pronounced that a strong case is made for collecting school data covering parochial and other private schools. The fact that Method II using public school data tends to give low estimates reflects the shift in recent years away from public schools and to parochial schools.

OTHER SOURCES OF ERROR IN METHOD II

Method II almost always requires the use of census statistics on children under 5 years old. In the census of 1940, as in most other censuses here and abroad, it is known that many children of this age were not enumerated. The census figures therefore have to be corrected for underenumeration of young children before the number of expected survivors is computed. Failure to do so would result in understatement of the expected population of school age and a commensurate decrease in the estimated net migration.

There are published measures of underenumeration for States, urban and rural; but for other areas one must select the measure that seems most appropriate from a priori considerations. There are, of course, intercounty variations in underenumeration within a State just as

there are in the completeness of birth registration. The variation is probably greatest in the States where underenumeration is greatest. For 1940 the estimated completeness of enumeration ranged from 79.9 per cent for South Carolina nonwhites to 98.0 per cent for Nebraska and South Dakota whites.[7]

If one is dealing with areas smaller than States, it usually will be necessary to distribute some age groups by single years of age since the single-year data were not published. This step can be performed fairly well on the basis of expected survivors of resident births during the appropriate years preceding the census. Even in areas with heavy migration, only minor errors should be introduced by this assumption.

Another point at which relatively minor errors may occur is in "aging" the corrected population distributed by single years to the estimate date.

A life table for the given area covering the given time usually will not be available. Hence the most appropriate life table must be selected. In some cases, it may seem desirable to adapt an existing life table. Recent life tables published by the federal government cover: (a) regions, urban and rural, by color and sex, for 1939;[8] (b) geographic divisions, by color and sex, for 1930 to 1939;[9] and (c) the white population of States, by sex, for 1939 to 1941.[10]

Although there is a fair amount of variation among areas in infant and child mortality rates, there is not much variation percentagewise in their complement, survival rates. For instance, according to 1929–1931 life tables for white males, by States, the proportion of the cohort under 7 years old surviving to be 7 to 13 years old varies only from .976 for Colorado to .986 for South Dakota. Nonetheless, it would be best to construct a life table for the area concerned, if it is large enough to have stable mortality rates. Actual deaths should not be subtracted from the population of the cohort at the base date since they include those of in-migrants and exclude those of out-migrants. It will be recalled that we are estimating the expected survivors assuming no migration.

Still another source of error arises from the fact that the ratio between the population of elementary school age and elementary school enrollment for any area is not constant over time. For States the

---

[7] U. S. Bureau of the Census. *Population*. "Differential fertility: 1940 and 1910. Standardized fertility and reproduction rates." p. 33. Washington, Government Printing Office, 1944.

[8] U. S. Bureau of the Census. "U. S. Abridged Life Tables, 1939, Urban and Rural, by Regions, Color, and Sex." June 23, 1943.

[9] U. S. Bureau of the Census. "U. S. Abridged Life Tables, 1930–39 (Preliminary), by Geographic Divisions, Color, and Sex." April 30, 1943.

[10] National Office of Vital Statistics. "State and Regional Life Tables, 1939–41." Washington, Government Printing Office. 1948.

average change in the ratio of the population 7 to 13 years old to public elementary school enrollment was +0.038 between 1930 and 1940. Disregarding sign, the average change was 0.047. The range was from −0.052 to +0.141. It has been shown in Table 2, however, that during the 'forties more accurate population estimates are obtained on the average by assuming that the 1940 ratio did not change than by extrapolating the 1930–1940 trend in this ratio.

We come now to what is probably the major assumption in our method. This is that the net rate of change through migration for elementary school-age children is equal to the net migration rate for the population of all ages. Migration is quite selective of certain age groups, of course; but there is some evidence that children of elementary school age, who move with their parents, are of intermediate mobility between young adults and the aged. Insofar as child migration is near the average rate, it may be used to represent the rate for the whole population.

Past experience in the pattern of net migration rates by age may be studied. From 1935 to 1940, the net interstate migration of children 5 to 13 years old was in the same direction as that of the total population in all except four States and the District of Columbia. Among the 44 States where net migration was in the same direction, the rate for children 5 to 13 years old averaged 91 per cent of that for all ages. The range was from 40 per cent to 167 per cent, with a standard deviation of 35 percentage points.

Among the 90 cities of 100,000 or more in both 1940 and 1930, the net migration of children 5 to 13 years of age was in the same direction as that of the total population in all except 5 cases. In these 85, the rate for the children averaged 145 per cent of that for all ages. The range was from 3 per cent to 400 per cent, the standard deviation being 69 percentage points.

From these facts, it can be appreciated that there is a great deal of variation in the relationship between the child-migration rate and the rate for the population of all ages. It is thus easy to make a large error in estimating the total net migration from the migration at the age of school enrollment. We can, of course, use the ratio observed in the particular State or city with which we are dealing. We do not know, however, how representative the ratio for 1935 to 1940 is for subsequent periods. There are still other disturbing factors. One is that in making current postcensal estimates we have to deal with periods ranging from one to nine years instead of just five years. Another is that the age group chosen to represent school enrollment may not be 5 to 13; for

instance it may be 7 to 14. Late in the decade, the school-age cohort includes some children born since the last census. This part of the cohort has had a shorter period of exposure to "risk" of migration than have persons of older ages. Fortunately, there is never a very large difference in the average length of exposure between the whole school-age cohort and the population of all ages.

A great deal of investigation remains to be done on the age selectivity of net migration for different areas over various periods of time. So far we have relatively few statistics on net migration by age. If we have taken care of the other sources of error in the method so that the migration assumption is the only important item of guesswork remaining, we can do some experimentation by trial and error. We can see what migration assumptions produce the most accurate population estimates, using a complete census or other reliable figure as the standard.

The foregoing discussion of sources of error was concerned with the estimation of net migration. The estimation of the actual natural increase is beset with relatively minor pitfalls. The available figures on under-registration of births are subject to some error. It is usually assumed that deaths are almost completely registered. There may also be some inaccuracy in the allocation of births and deaths to the correct area of usual residence.

The Census Bureau's State estimates, which are prepared essentially by Method II, are adjusted to total to a national estimate, which we consider to be quite comparable to a decennial enumeration. Similarly, county estimates made locally may be scaled up or down to a State total. Such controls reduce the chance of very large errors. Whenever censuses or better estimates are available for particular local areas, they should, of course, be substituted. For example, the method we have described proved inappropriate for Washington, D. C., because of the unusual character of migration to this city. As an experiment, this Bureau used a wide variety of statistics in preparing the estimates recently published.[11] We feel that although our school-data method is promising enough to warrant extensive experimentation and polishing, concurrent exploration of other sources and methods should be conducted.

[11] U. S. Bureau of the Census. *Population—Special Reports.* "Estimated population of the Washington, D. C., Metropolitan Counties: 1940 to 1946." Series P-47, No. 5. May 14, 1947.

THE USES AND USEFULNESS OF BINOMIAL PROBABILITY PAPER*

FREDERICK MOSTELLER

AND

JOHN W. TUKEY

Harvard University and Princeton University

This article describes certain uses of Binomial Probability Paper. This graph paper was designed to facilitate the employment of R. A. Fisher's inverse sine transformation for proportions. The transformation itself is designed to adjust binomially distributed data so that the variance will not depend on the true value of the proportion p, but only on the sample size n. In addition, binomial data so transformed more closely approximate normality than the raw data.

The usefulness of plotting binomial data in rectangular coordinates, using a square-root scale for the number observed in each category, was first pointed out by Fisher and Mather [10]. The graph paper under discussion[1] is specially ruled to make this mode of plotting both simple and rapid. A graduated quadrant makes the angular transformation ($p=\cos^2 \phi$ or $p=\sin^2 \phi$) easily available at the same time. Most tests of counted data can be made quickly, easily and with what is usually adequate accuracy with this paper. Some 22 examples are given.

PART I—GENERAL

INTRODUCTION

IN ANALYZING data which has been counted rather than measured the biologist, opinion pollster, market analyst, geologist, physicist, or statistician frequently uses as a model the binomial distribution, its limiting case the Poisson distribution, or some of their generalizations. For many purposes graphical accuracy is sufficient. The speed of graphical processes, and more especially the advantages of visual presentation in pointing out facts or clues which might otherwise be overlooked, make graphical analysis very valuable. A special type of graph paper has been designed and is now available for such analysis.

The examples below show how such paper can be used in various computations—clearly there are many parallel cases in other fields of analysis.

---

* Prepared in connection with research sponsored by the Office of Naval Research.

[1] Available from the Codex Book Company, 74 Broadway, Norwood, Mass. No. 31,298 on thin paper, No. 32,298 on thick paper.

The basic idea involved is due to R. A. Fisher, who, many years ago, observed that the transformation

$$\cos^2\phi_i = \frac{n_i}{n}$$

transformed the multinomial distribution with observed numbers $n_1, n_2, \cdots, n_k$ into direction angles $\phi_1, \phi_2, \cdots, \phi_k$ which were nearly normally distributed with individual variances nearly $1/4n$ (when the angles are measured in radians). Thus the point at a distance $\sqrt{n}$ from the origin and in the direction given by $\phi_1, \phi_2, \cdots, \phi_k$ is distributed on the $(k-1)$ dimensional sphere nearly normally, and with variance nearly $\frac{1}{4}$ *independent of n and the true fractions $p_1, p_2, \cdots, p_k$ of the different classes in the population.* The rectangular coordinates of this point are $\sqrt{n_1}, \sqrt{n_2}, \cdots, \sqrt{n_k}$.

In 1943 Fisher and Mather [10] applied this principle to the graphical plotting of the square roots of the observed numbers of *Lythrum salicaria* styles of various types on ordinary graph paper. After reading this article, the present authors were convinced that special graph paper, designed to facilitate this process would be worthwhile and designed the paper [12] mentioned above—whose use forms the subject of this article.

Binomial Probability Paper is graduated with a square-root scale on both axes 1(1)20(2)100(5)400(10)600 and 1(1)20(2)100(5)300. (The notation $A(a)B(b)C$ means: "From A to B by intervals of size a and thence from B to C by intervals of size b".) It serves to treat classification into two categories directly. Cases with three categories would require three-dimensional graphing for direct treatment, but a device not explained in this paper enables one to treat any number of categories on this paper.

PLOTTING

The binomial distribution

The classical example of the binomial distribution is the distribution of the number of heads in n independent throws of a coin, true (50-50), or biased (for example 40-60 or 57-43). Two assumptions require emphasis here (i) the throws are independent and do not influence one another, (ii) the probability of a head on each throw is the same. Replace "heads" and "tails" by "successes" and "failures," or by "regular females" and "exceptional females," or by "alcoholic" and "nonalcoholic," or by any classification into two categories and these two

requirements, suitably modified, are the requisite conditions that the observed numbers in the categories will be distributed binomially.

We must distinguish (i) the true relative probability of two categories, which we will often refer to as the *split* and (ii) the observed numbers in the two categories, which we will often refer to as the *paired count*. In describing a split we do not bother with normalization —the split of a "true" coin is 50-50 or 1-1 or 127-127. In describing a paired count we always give the actual numbers counted.

Binomial plotting

A *split* is plotted as a straight line through the origin—it passes through all the points whose coordinates represent it. Thus in Figure 1 the lowest line corresponds to the 80-20 split. In addition to going through (80, 20) this line goes through the points (8, 2), (16, 4), (20, 5), (120, 30) and so on—the corresponding splits are all the same. The 77-23 split is also called the 23% line, and the (100-X)-X split is also called the X% line.

A *paired count* is plotted as a triangle whose sides extend for one scale unit in the positive directions of the axes from its right angle, which has the paired count for its coordinates. Thus (1, 3) is plotted as the triangle with vertices (1, 3), (2, 3) and (1, 4). The paired counts (1, 3), (14, 1), (155, 4), (1, 110), (125, 150) are plotted in Figure 1. (The modification of the classical angular transformation thus induced is discussed in Part V (page 206).)

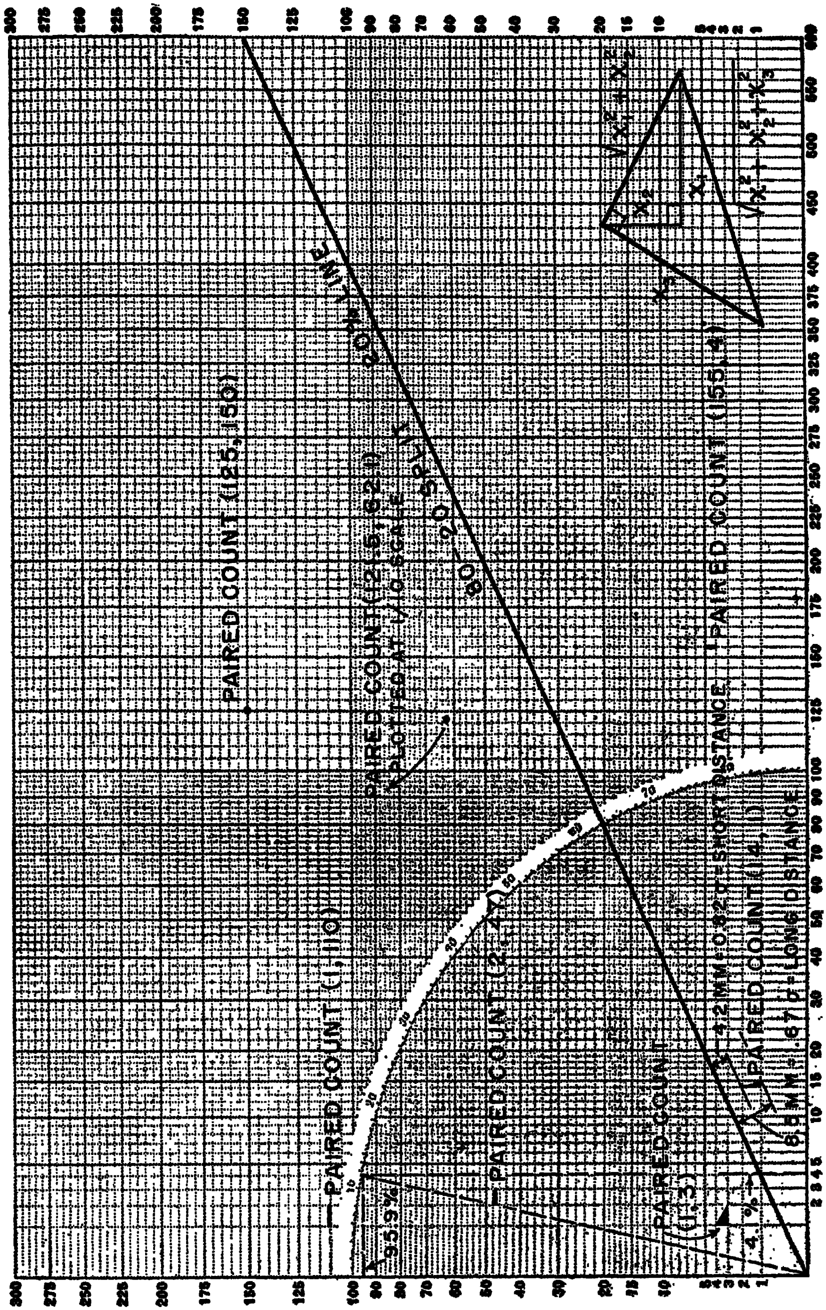
If any coordinate exceeds 100, it is usually satisfactory to plot a line segment (vertical in the case of (155, 4) and horizontal in the case of (1, 110)). If both coordinates exceed 100, it is usually satisfactory to plot a point.

The middle of a paired count is the center of the hypotenuse, which degenerates into the center of the line segment.

Percentages

1. *To convert a split into a theoretical percentage read the coordinates of its intersection with the quarter circle* (paying no attention to the angular scale).

2. *To convert a paired count into an observed percentage, draw the split through its lower left corner and read the coordinates where the split crosses the quarter circle.* (It is convenient to consider this an auxiliary line and draw it dashed). Thus (2, 47) corresponds to percentages of 4.1 and 95.9 (Figure 1).


PAIRED COUNT (125, 150)
PAIRED COUNT (215, 62.1)
PLOTTED AT 1/10 SCALE
20% LINE
80—20 SPLIT
PAIRED COUNT (155, 4)
PAIRED COUNT (1, 110)
PAIRED COUNT (2, 4)
PAIRED COUNT (1, 3)
PAIRED COUNT (14, 1)
4.2 MM = 0.320 σ = SHORT DISTANCE
8.5 MM = .670 σ = LONG DISTANCE
95.9%
4.1%

Measurements of deviations

A right triangle has three corners, but we shall only be concerned here with the two acute-angled ones. There are two distances from a line to these corners of a triangle. We call these the *short distance*, and the *long distance*. Thus, in Figure 1, the distances from the 80–20 split to the triangle representing (14,1) are 4.2 mm. and 8.5 mm. The distance to the middle of the hypotenuse, which is the average of the other two, is the *middle distance*, here 6.4 mm.

In some cases there are definite reasons, theoretical or empirical, to prefer the short distance, the long distance or a combination of the two. In the remaining cases it seems plausible to use the middle distance, though investigation may prove otherwise. In the examples below we have followed this plausible procedure.

Such distances are usually to be interpreted as nearly normal deviates. The scale in the upper left marked "Full scale—Individual Standard Errors" allows these distances to be read as multiples of the standard deviation.

For most purposes, however, a millimeter scale is more convenient. On the paper *one standard deviation is almost exactly 5 mm.* (if the paper has not shrunk or stretched, it is 5.080 mm.). Thus the short, middle and long distances of 4.2, 6.4 and 8.5 mm., from the example above, are 0.82, 1.24 and 1.67 standard deviations. A short table of the normal distribution is given in Part VI, both in millimeters and in standard deviations. For many purposes it is only necessary to remember that one deviation in three should be outside −5 mm. to 5 mm. and one in twenty outside −10 mm. to 10 mm. purely due to sampling fluctuations.

The reader who is interested in the simpler applications should now go to Part II (page 182). The reader who is interested in a specific application is referred to Part VI (page 209) which provides an index and outline.

REFINEMENTS IN PLOTTING

Plotting larger counts

Paired counts up to (600, 300) can be plotted directly. If there are more observations, but less than (6,000, 3,000), the paired count may be plotted at 1/10 scale, thus (1,215, 621) would be plotted as (121.5, 62.1) as in Figure 1. In this case

(1) the left-hand upper scale should be replaced by the right-hand upper scale in converting distances into multiples of the standard deviation;

(2) distances in millimeters should be multiplied by 3.16 before using Tables 1 and 2;
(3) the triangle would extend from (121.5, 62.1) to (121.6, 62.1) and (121.5, 62.2) and would be completely indistinguishable from a point.

Very large counts

If significance tests are required for still larger samples, graphical accuracy is insufficient, and arithmetical methods are advised. A word to the wise is in order here, however. Almost never does it make sense to use exact binomial significance tests on such data—for the inevitable small deviations from the mathematical model of independence and constant split have piled up to such an extent that the binomial variability is deeply buried and unnoticeable. Graphical treatment of such large samples may still be worthwhile because it brings the results more vividly to the eye.

Plotting unsymmetrical counts

Occasionally the user may have data involving large counts of one kind and small counts of the other, where the total count may not be fixed. Such cases may be treated by choosing a suitable divisor, and using the divisor only for the category with the large counts. If this is done, the distance from paired count to split should be measured at right angles to the contracted axis and *not* at right angles to the split. Each triangle reduces to a segment extending one scale unit away from the divided axis. For good accuracy, it is necessary that each small count be less than 5% of the corresponding total count.

Poisson subcase

The Poisson distribution yields the subcase of this case where the total number is arbitrarily large and *fixed*. Binomial probability paper may still be used without difficulty, plotting the various observations on any chosen, fixed vertical line, thus using an unknown horizontal divisor.

Crab addition

In order to sum the squares of deviations in a chi-square, we may proceed either by measuring, squaring and adding, or by crab addition, using the theorem of Pythagoras. Crab addition is easily done by using a piece of tracing paper as follows:

i) Mark a small circle as the origin (0) on the tracing paper and set

it on the theoretical line at the point of intersection with the perpendicular dropped from the first sample point;

ii) Mark a point (1) on the tracing paper at the sample point—the distance between the tracing paper origin and the first sample point is the square root of the contribution of the first sample point to χ^2, call it χ_1^2.

iii) Line up the origin (0) and the point (1) on the tracing paper on the theoretical line with the point (1) at the base of the perpendicular from the second sample point to the theoretical line, and plot a point (2) on the tracing paper immediately over the second sample point. The distance from point (1) to point (2) is χ_2 the square root of the contribution of the second sample point to χ^2, and the distance from the origin (0) to point (2) is $\sqrt{\chi_1^2+\chi_2^2}$. We continue in this manner producing a picture like that shown in the lower right-hand corner of Figure 1.

Use for chi-square

If the individual segments submitted to crab addition were *middle distances*, then the final crab sum may be (geometrically) doubled in length and read on the marginal scale to obtain a χ^2 value. (Multiply this answer by another factor of 10 if the points were plotted to 1/10 scale.)

If the individual segments submitted to crab addition were *ranges* then the final crab sum may be read on the marginal scale and then (numerically) doubled in value to obtain a χ^2 value (see Ex. 24). (Multiply by another factor of 10 if the points were plotted to 1/10 scale.)

SIGNIFICANCE LEVELS AND SIGNIFICANCE ZONES

The continuous case

When a value of a continuously distributed statistic, such as Student's "*t*," is computed from a sample, there is no difficulty (provided the distribution is well tabled) in finding and stating its level of significance. Thus a t of -0.98 on 3 degrees of freedom lies at (i) the lower 20% point, (ii) the upper 80% point, (iii) a (balanced) two-sided 40% point. These statements mean that, *if* the simple situation *against* which we are assessing the evidence really holds, and similar experiments are repeated, (i) an algebraically smaller value (that is, <-0.98) of t will occur in 20% of the cases, (ii) an algebraically larger value (that is >-0.98) will occur in 80% of the cases, (iii) a value further from the center (that is, $|t|>0.98$) will occur in 40% of the cases. The two-sided statement is apparently simple to make in the case of t, since

the distribution concerned is symmetrical and continuous. In asymmetrical cases, we will follow the path of least resistance and calculate the two-sided significance level as twice the (smaller) one-sided significance level, just as 40% = 2(20%). (Some statisticians might disagree.) The choice of one among the three significance levels in a practical situation will depend on the alternatives considered for the situation against which the evidence is being assessed.

The discrete case

In the binomial situation, and in the many other cases where the result obtained proceeds in definite steps, the situation is a little more complex. Consider the case of a sample of 4 from a 2–1 split. The probabilities of the various possible outcomes are given below to 3 decimal places.

Outcome:	(0, 4)	(1, 3)	(2, 2)	(3, 1)	(4, 0)
Frequency:	.012	.099	.296	.395	.198
Cumulated from left to right:	.012	.111	.407	.802	1.000
Cumulated from right to left:	1.000	.988	.889	.593	.198

What significance level shall we assign to (1, 3) in this case ? The conventional answers are that it lies at: (i) the lower 11.1% point, (ii) the upper 98.8% point, (iii) the two-sided 22.2% point. These statements mean that, if the situation were binomial with a 2–1 split, (i) one of the outcomes (0, 4) or (1, 3) occurs in 11.1% of all cases, (ii) one of the outcomes (1, 3), (2, 2), (3, 1), (4, 0) occurs in 98.8% of all cases, (iii) it is reasonable to act as though an outcome deviating from expectation *as much or more* than (1, 3) in one direction or the other occurs in 22.2% of all cases.

An alternative approach, and one which supplies more information, is to attach to such a result as (1, 3) not a single significance level, but a *significance zone*. Thus the lower significance zone for (1, 3) is "11.1% to 1.2%" and the two-sided significance zone is 22.2% to 2.4% (where we have adopted the convention of doubling in passing from one-sided to two-sided significance zones.)

The statement "(1, 3) is at the lower (11.1%, 1.2%) zone" means that, if the simple binomial situation with split 2–1 holds, then (0, 4),

which is the only outcome further from expectation than (1, 3) *in the same direction,* will occur in 1.2% of all cases, while (0.4) and (1, 3) *together* will occur in 11.1% of all cases.

Interpretation

The working statistician, we believe, would almost always react differently to the statement—the outcome is at the lower (12%, 0.5%) zone—than to the statement—the outcome is at the lower (12%, 11.5%) zone. The (12%, 0.5%) zone indicates that we are not sure that strength of the evidence has reached the conventional 5% point, but it is possible that this is the case. The (12%, 11.5%) zone indicates that the strength of the evidence has certainly not reached the conventional 5% point. This distinction is absent from the customary procedure of assessing just a significance level.

In some cases, auxiliary experimental data could be used to interpolate between the ends of the significance zone.

As we shall see, binomial probability paper can be used to obtain both the customary significance level and the significance zone. If we are to determine the two percentages needed to fix a significance zone, we shall need to make two measurements on the paper, so it is not surprising that we will want to plot more than a single point to represent a paired count.

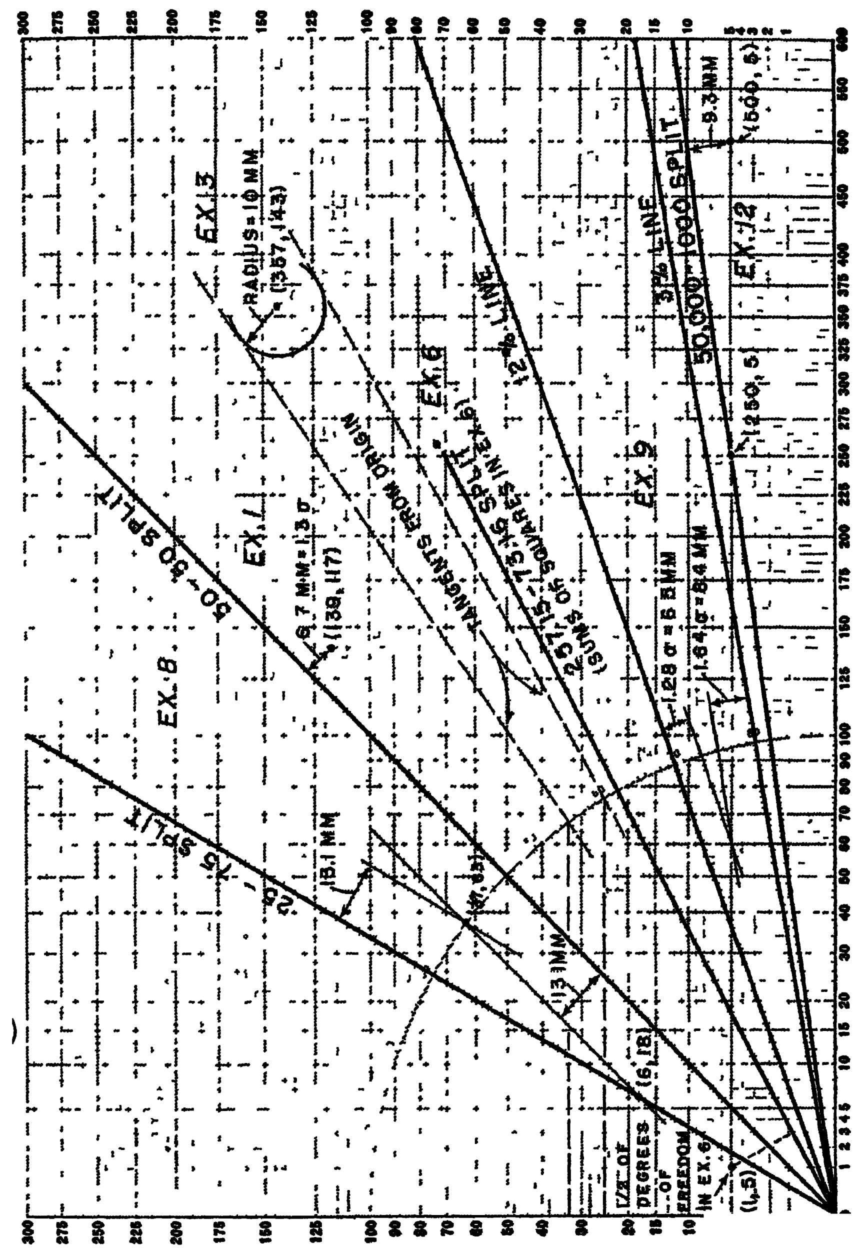
PART II—PLOTTING ONE OBSERVED QUANTITY

A SINGLE BINOMIAL SAMPLE

Given a sample, sorted into two categories, the numbers in the categories form a paired count, which is plotted as a triangle. If a probability or a population proportion, p, is assigned for the second category, then this is plotted as the q-p split, where $q=1-p$. The approximate significance level corresponds to the short distance from the split to the triangle, and may be obtained from Table 2 (page 211). If it is desired to use the significance zone, then both short and long distances should be measured and the probabilities obtained from Table 2. These general principles will suffice to attack the problems posed by the first four examples.

Comparing an observed proportion with a theoretical proportion

Example 1 (See Figure 2). Fisher and Mather [11] have described a genetical experiment in which the individuals in 32 litters of 8 mice were observed for the characteristic of straight versus wavy hair. Under the conditions of the experiment, the Mendelian theory



NUMBER—WITH WAVY HAIR (EX. 1), OF OWNED FARMS (EX 3), WITH CHARACTER ABSENT (EX. 8) OF DEFECTIVES (EX. 9), OF DUMMY SUCCESSES (EX 12).

predicted that half the mice would have straight hair. It was observed that $n_1=139$ had straight hair, $n-n_1=117$ had wavy hair. Could such a discrepancy from the 128-128 split have arisen by chance?

The observed paired count is plotted as the triangle (139, 117), (139, 118), (140, 117) and is just distinguishable from a point. The theoretical proportion is plotted as the 128-128 split (the 50% line). The short distance is about 6.7 mm. $=1.3\sigma$. Since deviations from simple Mendelian genetics might reasonably occur in either direction, the two-sided significance level is appropriate, and from Table 2 this is found to be about 20%. Thus this test would not lead us to reject simple Mendelian genetics.

Since the two-sided 5% distance is 10 millimeters, which is worth remembering, Table 2 would not have been needed in routine testing, since the result "not significant at the 5% point" would have been enough.

We may observe the experimental percentage of straight-haired mice by drawing a horizontal line to the vertical axis from the intersection of the degrees quarter-circle with the line from the observed point to the origin. This gives about 54.5 per cent as compared with an actual value of 54.3 per cent.

If the significance zone were desired, then both the short distance of 6.7 mm. and the long distance of 7.5 mm. would have been measured. The corresponding two-sided significance interval is (20%, 14%).

Critique of Example 1. The experimental conditions seem to have been exactly suited for a binomial test. The accuracy of the graphical method is clearly adequate.

The sign test

The classical case of the sign test is the comparison of two materials or treatments, in pairs, where the observations in each pair are comparable except for the materials or treatments being tested. The sign test is a special case of the comparison of theoretical and observed proportions, where the theoretical proportion is always 50% and has been thoroughly discussed in this Journal by Dixon and Mood [4, 1946].

Example 2 (See Figure 3). Dixon and Mood cite the yields of two lines of hybrid corn where 6, 8, 2, 4, 3, 3, and 2 pairs of plots were available from 7 experiments. In 7 out of the 28 pairs line A yielded higher.

If a significance zone is wanted, then both the short distance of 12.9 mm. and the long distance of 14.7 mm. are measured. The resulting two-sided significance zone is (1.2%, 0.3%).

Comment on Example 2. The tables of Dixon and Mood correctly state that (7, 21) is not at the 1% level. Some statisticians, however, in part because of the general and unprecise considerations which lead to the choice of 1%, may use the extra information in the significance zone and decide that they would rather work at the 1.3% level than the 0.4% level (the precise values are 1.254% and 0.372%). They would then treat (7, 21) as significant at a level of approximately 1%.

Extension of the sign test

As presented above, the sign test applies to the hypothesis of equality in paired experiments. It can be easily extended to cover (i) the hypothesis of a constant additive difference, (ii) the hypothesis of a constant percentage difference, or (iii) the hypothesis of a certain population median by constructing dummy observations. Suppose that five experiments have produced the following numbers:

Set	1	2	3	4	5
Condition A	57	53	49	56	51
Condition B	26	31	24	28	31

There is clearly no need to test for equality. A test of the hypothesis that condition A runs 30 units above condition B may be made by adding 30 to each observation under condition B, which yields 2 positive and 3 negative differences. A test of the hypothesis that condition A gives numbers double those of condition B may be made by doubling the condition B numbers, which yields 2 positive, 1 zero, and 2 negative differences (the corresponding paired count is (2, 2)!). A test of the hypothesis that the median number in condition A is 57, may be made by replacing the results in condition B by an imaginary experiment always giving 57, which yields 1 positive, 1 zero, and 3 negative differences (the corresponding paired count is (1, 3)!).

Confidence limits for population proportions

If we have a sample divided into two categories, we may wish to set confidence limits on the percentage of the population in each category. This is accomplished by plotting the paired count which was observed and then constructing two splits whose short distances from this triangle correspond to the two-sided level (of confidence) required. Thus if 95% confidence is required, short distances of 10 mm. will be used.

The coordinates of the intersections of the two splits with the quarter-circle give the confidence limits for percentages.

Example 3 (See Figure 2). A random sample of 500 farms from a certain region yields the information that 143 of the farmers own outright the farms they work. Set 95 per cent confidence limits on the per cent of farms in this region wholly owned by the farmers.

We plot the triangle (357, 143), (358, 143), (357, 144), which can, in this case, be approximated by a point. We draw arcs of circles of radius 10 mm. about its extreme vertices, and draw the tangent splits. We get as our estimate of the percentage of farmer-whole owners 28.5 (compared to 28.6 computed) and as our 95% confidence limits, 25 per cent and 33 per cent.

Critique of Example 3. If the sample is truly random the method is sound. If stratified sampling is correctly done, these confidence limits will be unnecessarily large by an amount which depends on the effectiveness of the strata chosen, for the particular question at hand. The increase in efficiency by stratification is often so small that these limits are a wise choice.

Confidence limits for the population median

By combining the ideas of the sign test and the last example, we may obtain confidence limits for the population median. If x_0 is the population median, and is used to divide samples into paired counts, then certain extreme paired counts will occur with probability at most $1-\alpha$, where α is the desired confidence. By determining these unlikely splits, and referring to the observations, it is possible to set confidence limits for the population median. While strict confidence requires taking blocks of paired counts with at most a certain probability, interpolation can usually be carried out with safety.

Example 4 (See Figure 3). The differences between reaction-times of an individual and a control group were 6, 5, 3, 2, 1, -1, -2, -3, -5, -12, -13, -13, -15, -28, when expressed on a logarithmic scale (actually in terms of 100 times the log to the base 10). Within what limits do we have 90% confidence that the median of the population lies?

Drawing the 50–50 split, and parallel lines at ± 8.4 mm., we find that the triangles for (10, 4) and (4, 10), where $4+10=10+4=14$ (the number of cases), are cut by these parallel lines about $\frac{3}{4}$ of the way from their vertices nearest the 50–50 split. We interpolate $\frac{3}{4}$ of the way from 1 to 2 (these are the fourth and fifth values from the top and $\frac{3}{4}$ of the way from -12 to -13 (these are the fourth and fifth values from the

bottom) to obtain approximate 90% confidence limits of $-12\frac{3}{4}$ and $1\frac{3}{4}$ for the median of the population of differences from which the 14 observed differences were a random sample.

Critique of Example 4. The interpolation is based on the fact that the use of 1.001 (say) for a cutting point would give the paired count (4, 10), as would any cut up to 1.999, while 2.001 would give (3, 10). If we take account of the grouping and rounding process we should widen these interpolated values by $\frac{1}{2}$ the grouping interval. Thus if the difference were rounded from more decimals to the nearest tenth, and were actually 6.0, 5.0, 3.0, $\cdots$, -28.0 then we should use $-12\frac{3}{4}-\frac{1}{20}=-12.8$ and $1\frac{3}{4}-\frac{1}{20}=1.8$. If, as might more reasonably have been the case, they had been rounded to integers, we should use $-12\frac{3}{4}-\frac{1}{2}=-13.25$ and $1\frac{3}{4}+\frac{1}{2}=2.25$.

COMPARISON AND ANALYSIS OF VARIANCE

Comparison

The comparison of two variances calculated for samples from normal populations, leads alternatively to Fisher's z or Snedecor's F. As explained in Section 14, the distribution of these quantities is mathematically related to the binomial distribution. Thus we may use binomial probability paper to make a significance test of the equality of the variances of two normal populations—the only approximation being the approximation of binomial probability paper to the binomial distribution.

The test is made by drawing the *line* or *split* through the point whose coordinates are the observed *sums of squares* of deviations from means, and plotting the *point* whose coordinates are half the numbers of degrees of freedom.

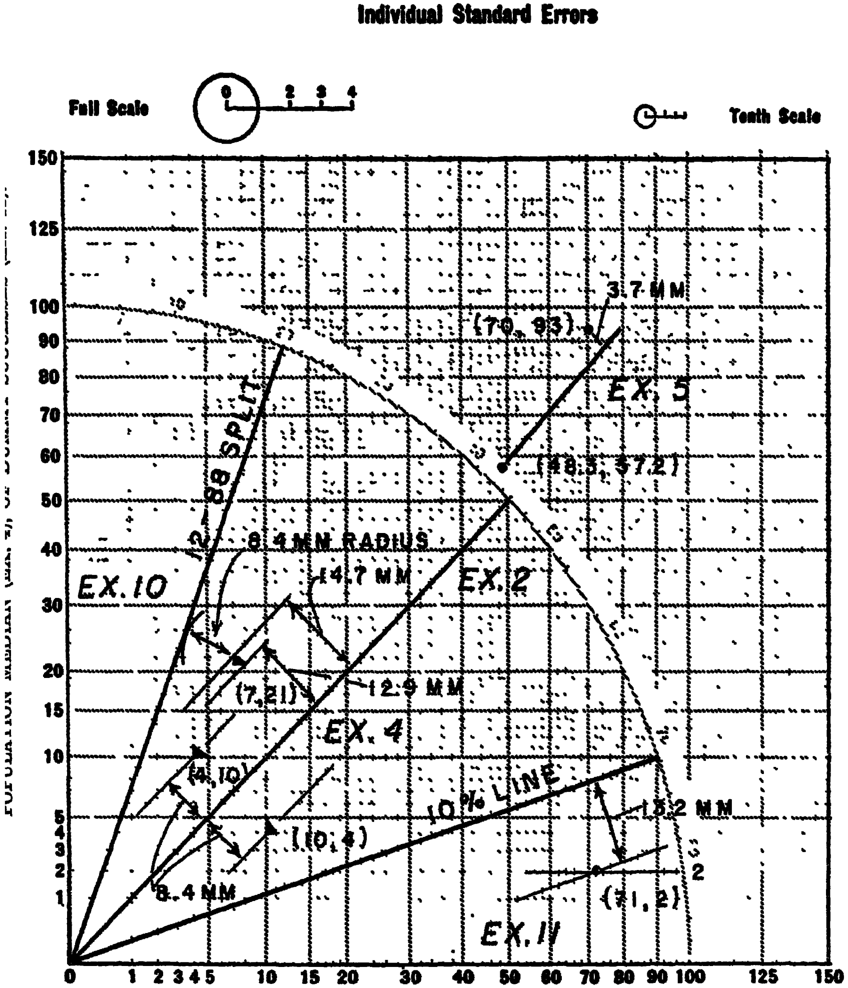
Example 5 (See Figure 3). In volume 1 of Biometrika, Fawcett [8, 1902, p. 442] gives the sample variances of the lengths of 141 male Egyptian skulls as 34.2740 and the variance of the lengths of 187 female skulls as 30.5756. Are these significantly different?

The sums of squares of deviations are 4832.6 and 5717.6. The degrees of freedom are 140 and 186. Drawing the split through (48.3, 57.2) and plotting the point ($70=140/2$, $93=186/2$), we find a distance of 3.7 mm which is near a (two-sided) significance level of 45%. Thus there is no evidence of a difference in variance.

Critique of Example 5. The dangers in such a test lie far more in the possible lack of normality of the populations of practical experience than in the approximations of binomial probability paper. Such comparisons of variances from *independent* samples based on normality

FIGURE 3

SHOWING THE SIGN TEST (EX. 2), CONFIDENCE LIMITS FOR MEDIAN (EX. 4), THE F TEST (EX. 5), OPERATING CHARACTERISTIC OF SIGN TEST (EX. 10) SAMPLE SIZE FOR POPULATION TOLERANCE INTERVALS (EX. 11)



NUMBER—OF PAIRS WITH LINE A HIGHER (EX. 2, 10), OF CASES BELOW POPULATION MEDIAN (EX. 4), OF DUMMY FAILURES (EX. 11)

should almost always be taken with a grain of salt, particularly when so many degrees of freedom are involved.

This procedure is quite convenient when both degrees of freedom are greater than 24 (this is outside the range of short tables of F or z) and will be quite accurate if both degrees of freedom are at least 2. When one of the degrees of freedom is 1, accuracy will be improved by plotting $\frac{1}{2}$ seven-tenths of the way from zero to one, because $\sqrt{.5}=.7$. The use of binomial probability paper has the feature, interesting to some, of producing estimates of intermediate levels of significance. If it is desired to check the tabled values of levels of F against the results given by binomial probability paper, one plots the split (n_1F, n_2) and measures the distance to the point $(n_1/2, n_2/2)$, remembering that the tabled values provide a one-sided test.

Analysis

The same process can, of course, be applied to the analysis of variance, in principle a special case of the comparison of variances, but in practice a whole realm of its own. The sum of squares column determines the line and the *df* column the point.

Example 6 (See Figure 2). How can the analysis of variance in Example 7 below be tested graphically?

Draw the line through (257.15, 73.16) and plot the point (1, 5). The perpendicular distance of 15 mm is very highly significant (beyond the 0.1 per cent point).

Critique of Example 6. The difficulties with normality, mentioned in the last critique, are usually reduced enough to be neglected in an analysis of variance situation. Only in case this graphical method yields borderline significance is accuracy an excuse for using an F or z table, though convenience may often be a reason.

The angular transformation

The analysis of variance of counted data is frequently facilitated [3, Bartlett 1947 and references cited therein] by making the angular transformation. If the data involve small numbers, the accuracy of graphical transformation will suffice. The observation is plotted, a line through the origin produced to the quarter-circle, and the corresponding angle read off.

Example 7. W. E. Kappauf and W. N. Smith (personal communication) tested the performance of six observers in reading three types of dials. Sixty readings were made by each observer on each size. The

errors, angles, and analysis of variance are shown below. (The reader will find it easy to check on the computation of the angles on his own piece of graph paper.)

	Errors in 60 trials			Corresponding angles		
Observer	Size I	Size II	Size III	Size I	Size II	Size III
B_0	7	4	4	20.2	15.0	15.0
B	15	6	3	30.2	18.6	13.1
K	10	7	5	24.1	20.2	16.9
R	28	20	16	43.2	35.6	31.2
S	11	8	9	25.2	21.6	23.0
T	25	17	15	40.0	32.2	30.2

Source	df	Sum of squares	Mean square
Observers	5	991.64	198
Sizes	2	257.15	129
Interaction	10	73.16	7.3
Binomial	—	—	$\frac{821}{60}=13.7$*

* Variance of an angle obtained from binomially distributed data $\frac{1}{4n}$ (radians)^2$=\frac{821}{n}$ (degrees)^2.

The usual test for significance of the effect of size would be $F=129/7.3=17.8$ on 2 and 10 degrees of freedom, which is highly significant.

Critique of Example 7. A possibly more conservative test would be $F=129/13.7=9.4$ on 2 and ∞ degrees of freedom, which is still very highly significant. Although 7.3 on 10 *df* is not significantly less than 13.7, there is reason to believe in this case that the error mean square in a large-scale repetition of this experiment might well be less than 13.7. For the analysis above assumes the errors distributed binomially, and the probable differences in difficulty of the various dials attempted might reduce this variance.

PART III. APPLICATIONS TO DESIGN

BINOMIAL DESIGN

Since binomial probability paper allows us to approximately judge the significance of a paired count, it must also let us plan binomial experiments to have desired properties.

Sample size necessary to resolve two given percentages

When designing a test to discriminate between two theoretical percentages, the experimenter often wishes to know that *any* result will give significant evidence *against at least one of the two theories.* The procedure is best described by an example.

Example 8 (See Figure 2). A geneticist wishes to test whether a certain character appears in one-half or one-quarter of the progeny of a certain mating. He requires significance at the (two-sided) 1 per cent level against at least one of these hypotheses and wishes to know the smallest sample size which will guarantee this. He draws the 50-50 and 25-75 splits and then parallel lines at a distance of 13.1 mm (2.58σ) which corresponds to the two-sided 1% level. These parallel lines intersect at (37, 63). This point separates the triangle (36, 63), (37, 63), (36, 64) from the triangle (37, 62), (38, 62), (37, 63). Thus the paired count (36, 63) is beyond the 1% level from the 50-50 split, while (37, 62) is beyond the 1% level from 25-75. Thus a sample size of $36+63=37+62=99$ ($=37+63-1$) will be enough.

Critique of Example 8. The design of this experiment is very good as far as sample size and significance levels are concerned. Since such genetical ratios are usually well-behaved, the experimenter who uses a sample of at least 99 and protects the progeny against causes of differential mortality should obtain very good results.

Our criticism should be directed against his less careful competitor who says that he will use the two-sided 1% level also, but will be satisfied with a sample size for which an observed 75-25 proportion will be at this level. Since the parallel line for 50-50 cuts the 25-75 split at (6, 18), he will use a sample size of only $5+18=6+17=23$. This design probably does not meet his needs, for if he uses it over and over he will be well protected from falsely stating the ratio is not 50-50 (since he is using a two-sided 1% level) *but he will miss one-half of the cases where the ratio is* 25-75. Such designs emphasizing one risk are usually ill-chosen, and if such a choice is compelled by limited experimental resources the choice of significance level should be re-examined, looking at both types of risk.

Designing single sampling plans

An essentially equivalent problem arising in industrial work as well as in certain kinds of experimental work is to design a sampling inspection plan which will distinguish between two kinds of quality, say product which is $100p_1$ per cent defective and that which is $100p_2$ per cent defective ($p_1 < p_2$). The plan desired is often described in terms of the operating characteristic curve, namely that large lots having $100p_1$ per cent defectives should be accepted $100(1-\alpha)$ per cent of the time, while lots having $100p_2$ per cent defectives should only be accepted 100β per cent of the time ($1-\alpha > \beta$). The process of building such a plan is described by the following:

Example 9 (See Figure 2). It is desired to construct a plan which will accept product with 3 per cent defectives, 95 per cent of the time, while product 12 per cent defective is to be accepted only 10 per cent of the time. What sample size should be used, and how many defectives can be tolerated before the lot is rejected?

We construct the 3% line (the 97-3 split) and the 12% line (the 88-12 split). If we wish to accept lots which are 3 per cent defective 95 per cent of the time, we must accept lots whose samples, as plotted on the paper, go as high as 1.64σ above the 3 per cent line, consequently we draw a parallel line $1.64\sigma = 8.4$ mm above the 3% line. Similarly in order to reject material which is 12 per cent defective 90 per cent of the time, we must reject lots whose samples come within 1.28σ of the 12% line, so we draw a parallel line $1.28\sigma = 6.5$ mm below the 12% line. The intersection of the two construction lines is (61, 5) and we find that the sample size is $61+5-1=65$, the acceptance number is 4, the rejection number is 5.

Critique of Example 9. It is assumed here that the lot is much larger than the sample, say ten times as large. Notice that these significance levels (10% and 95%) were one-sided. Consultation of tables shows that this plan will accept 3%-defective lots 96.77% of the time and will accept 12%-defective lots 9.69% of the time, which is approximately the result requested.

The operating characteristic of the sign test

We often want to know how well the sign test will discriminate.

Example 10 (See Figure 3). In Example 2 we considered 28 pairs of observations and decided to treat (7, 21) as significant. It is natural to inquire what population percentage of favorable pairs is needed to insure significance at this level 95% of the time. We must then find a

split so that the triangle representing (7, 21) is 8.4 mm $=1.65\sigma$ away. This leads to the 12-88 split, and so the sign test with 28 pairs discriminates very well between 50% and 12% (or 88%).

As in Examples 8 and 9, we can determine a sample size so that the sign test will have given discriminating power.

TOLERANCE LIMIT DESIGN

Sample sizes for population tolerance limits

In industrial work it may be desirable to take the least sample from an unknown population, such that the range from the smallest value in the sample to the largest value in the sample will cover a given fraction a of the population with given confidence β. This may be shown to be equivalent to a binomial problem, namely: Find the least sample size from a q-p split ($p=1$-q) such that the second count will be at least 2 with confidence β. (This is, of course, a special case of Example 8.)

If it is desired to use the rth from the bottom and the mth from the top to establish tolerance limits, replace 2 by $r+m$.

Example 11 (See Figure 3). A manufacturer of ball bearings wishes to have 99.5% confidence that 90% of his ball bearings lie between the limits set by the largest and smallest of a sample of a chosen size. He draws the 10% line (10% = 100% − 90%!), a parallel line $2.58\sigma=13.2$ mm lower (for 99.5% confidence) and the horizontal line through 2. The intersection of the last two lines gives (71, 2) and the desired sample size is $74=72+2$.

Critique of Example 11. This example assumes that the successive ball bearings produced behave like a random sample—no manufacturer of any metal object, even ball bearings has reached so high a state of control. The practical interpretation of such a sample size is that it is a lower bound.

Second sample tolerance limits

When tolerance limits are desired for a second sample, rather than for the population, the problem is hypergeometric, and may only be approximated by a binomial problem. The chance that the range between the rth from the bottom and the mth from the top from a sample of n will omit N_1 or less from a second sample of N, may be shown to be the same as that a sample of N_1+n_1 (where $n_1=r+m$) from a finite population split N to n will contain n_1 or more of the second sort. If $N_1 \leq N/10$, this can be approximated by the chance

of a second count of n_1 or more in a sample of N_1+n_1 from a N-n split.

Example 12 (See Figure 2). A manufacturer of precision resistors tests random samples of 1000 of each new type, and establishes the second from each end as working limits. He wishes to know the confidence with which he may expect (a) 99.5 or (b) 99 per cent of a batch of 50,000 similar resistors to fall within these working limits. He draws the 50,000-1,000 split and computes the chances of getting a second count of 4 or more in a sample of (a) 250+4, (b) 500+4. These are given by the distances from the split to (250, 5) and (500, 5), which are 0.0 mm≈50.0% and 9.3 mm≈3.4%. He has, therefore, 50.0% confidence that 99.5% will be within working limits and 96.6% confidence that 99% will be within working limits.

In the light of this example, the reader can construct answers to the other problems of tolerance limits for a second sample.

Critique of Example 12. If the manufacturer's production line is nearly in control when it starts to produce a new type, the authors will be surprised. The procedure given will answer the manufacturer who "wishes limits to the confidence, PROVIDED the process were in perfect control from the start." Also note that m and r have to be selected before the sample is examined, if the probabilities are to be accurate.

ANALYSIS OF VARIANCE DESIGN

Operating characteristic of model II—anova

An analysis of variance situation is Model II [6, Eisenhart 1947], when the effects are drawn from a normal population with variance σ_1^2, the errors being drawn from a normal population with variance σ^2. If the effects are the c column effects in a simple design with c rows and r columns, then the mean squares for columns and for error have the expectations $\sigma^2+r\sigma_1^2$ and σ^2 and the degrees of freedom (c-1) and (c-1) (r-1). In Model II, the mean squares are still distributed like multiples of chi-square. Thus the power of such an experiment can be easily determined as in the following example.

Example 13 (See Figure 5). A random sample of 25 sailors are to have their balancing ability measured quantitatively on each of 7 days. The results will be submitted to the analysis of variance, and a 5% significance level used. How large must σ_1^2/σ^2 be, before the existence of differences between sailors will be detected 95% of the time?

Here $c=25$, $r=7$, and the degrees of freedom are 24 and 144, hence we plot the point (12, 72). A one-sided 5% level corresponds to 8.4

mm, so we draw a circle around (12, 72) with this radius. The tangent splits are 27.2-100 and 8.9-100, so that the critical variance ratio is found from

$$\frac{\sigma^2 + 7\sigma^2}{\sigma^2} = \frac{27.2}{8.9}$$

to be $\sigma_1^2/\sigma^2 = .29$.

The basis of this construction is as follows: A ratio of sums of squares of 27.2 to 100 is needed for the chosen significance level. A ratio as small as 8.9 $(\sigma^2+7\sigma_1^2)$ to $100\sigma^2$ can be expected by chance 5% of the time if the population ratio is $\sigma^2+7\sigma_1^2$ to σ^2, hence to obtain 95% confidence of finding significance at 5%, we must have

$$\frac{8.9(\sigma^2 + 7\sigma_1^2)}{100\sigma^2} \geqq \frac{27.2}{100}.$$

PART IV—SEVERAL PAIRED COUNTS

LEVEL AND HOMOGENEITY COMBINED

Comparing several sets of data with a theoretical proportion

A set of several paired counts may give evidence against a fixed theoretical proportion in two ways. If the observed proportions are too variable, they indicate lack of homogeneity, that the samples came from populations with different proportions. If the average observed proportion deviates too much from the theoretical, it indicates a change (or error) in level, that the theoretical proportion does not apply.

It is well known that the correct way to test a homogeneous set of paired counts for agreement with a preassigned population proportion is to add them together, and then test the sum (as in Part II). Tests of homogeneity alone are the subject of the next section. Many delicate testing procedures involve first a test of homogeneity and then a test of level. Combined tests are mainly used to make quick and easy tests.

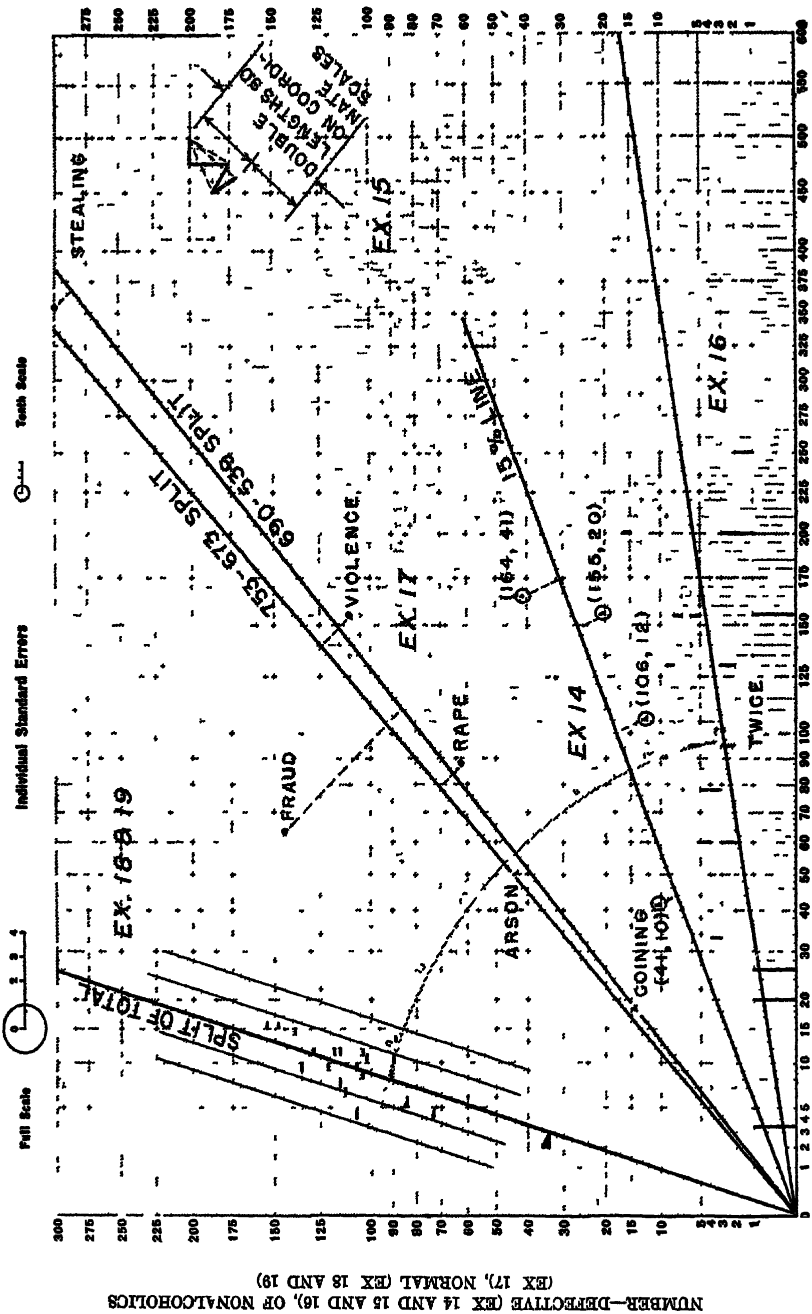
Example 14 (See Figure 4). A production process has been producing an average of 15% defective pieces over a long period. After the introduction of a new batch of raw material, successive shifts produced the following paired counts of nondefectives and defectives:

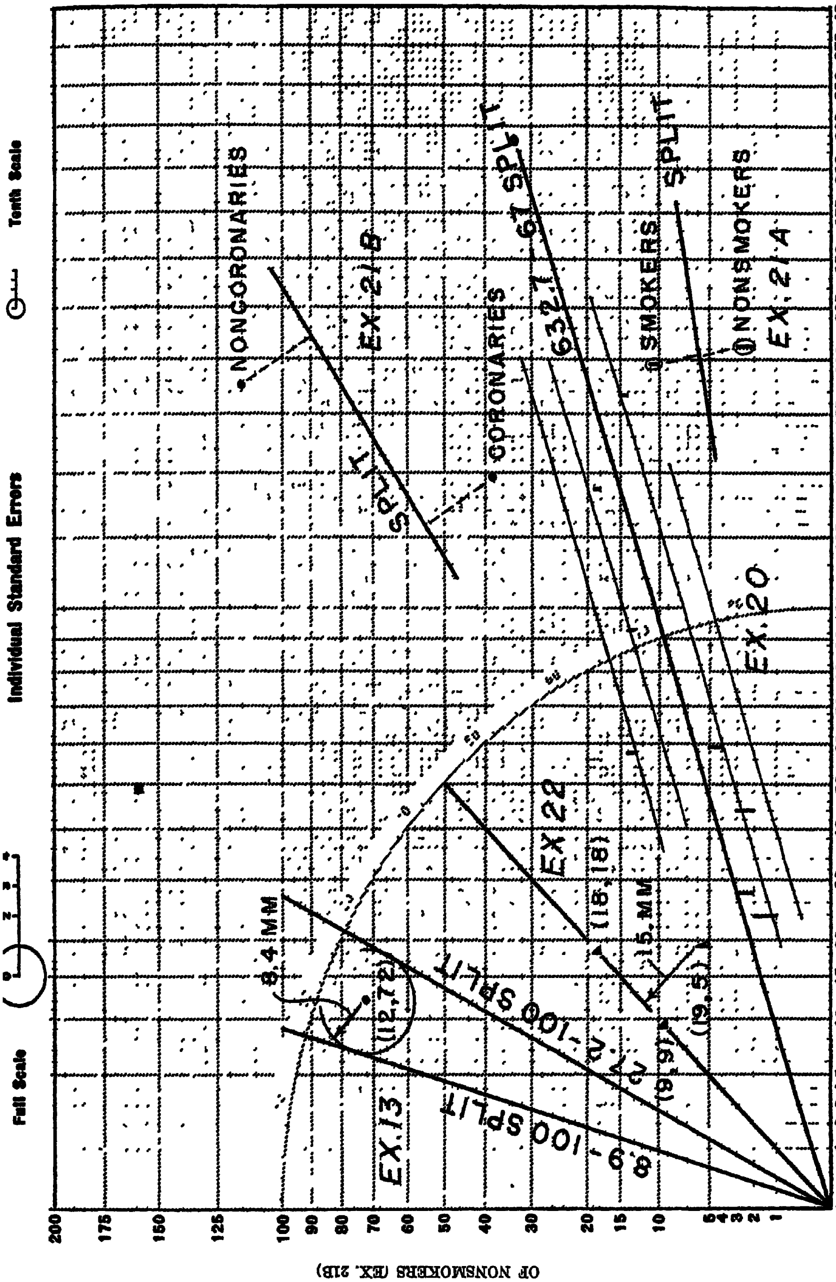
1. (155, 20), 2. (164, 41), 3. (106, 12), 4. (41, 10).

Is it reasonable to think that the production process is producing the same proportion of defectives as before? We plot the points and the

FIGURE 4

SHOWING COMPARISON OF K PROPORTIONS WITH THEORY (EX 12, EX 15), FIRST STEP TOWARD A STABILIZED p CHART (EX 16), HOMOGENEITY OF K PROPORTIONS (EX 17, EX 18 EX 19)





NUMBER—OF DEGREES OF FREEDOM ($X\frac{1}{2}$) (EX. 18), SHOWING PRIMARY NONDISJUNCTION (EX. 20), OF CORONARIES (EX. 21A), OF NONSMOKERS (EX. 21B)

15% line. Since no middle distance from the line is as great as $2\sigma=10$ mm, and since the points are about equally spread about the line, there seems to be no reason to suppose the new batch of raw material has made a change in the percentage defective.

If some points were found outside 2σ the resulting paired count (of points inside 2σ vs. points outside 2σ) would be compared with the 19-1 split by the method of Example 2.

An alternative procedure is to combine the middle distances from the theoretical line by crab addition (on page 179) and consider this a chi-square with as many degrees of freedom as there are paired counts.

Example 15 (See Figure 4). Taking the same data as in Example 14, the crab sum, which can be obtained in less time than it takes to read the description, is a length, which when doubled may be read on the marginal scale as $\chi^2=9.0$ which is between the upper 10% and upper 5% points for 4 degrees of freedom.

Critique of Examples 14 and 15. Merely examining to see whether any points fall outside the 2σ limits does not squeeze the data dry, and is not very precise, but is very convenient in situations where a control chart would seem reasonable and proper. Both tests tend to detect either a change to a new level or excess variability due to changes in level from shift to shift. To check one of these alone, different procedures should be used. A change to a new level should be tested for by combining all the shifts, and thus comparing (466, 83) with 85-15 (which is far from significance). Changes in level from shift to shift should be tested for by the methods of the next examples (Ex. 17–19).

The stabilized p-chart

Where the lot size is constant in 100% inspected industrial production or the sample size is constant when sampling inspection is used, one of the standard quality control procedures is the p-chart, where the percentage of each lot or sample found to be defective is plotted against lot number or time.

When lot size is not constant, and it is not feasible to break down the data into groups of different size, there is a need for a new technique. The use of groups of varying size is not recommended—it would be better to break up the lots into rational sub-lots of nearly uniform size—but where the quality control engineer cannot arrange for the better solution he may wish to use the following device which was suggested to us by Acheson Duncan [5]. (The classical device is

to plot observed percentages, with broken horizontal lines for control limits. This makes a hard-to-read, messy diagram.)

Example 16 (See Figures 4 and 6). The following data on adjustment irregularities of electrical apparatus appear in the ASTM *Manual on Presentation of Data* [1, 1940 *Supplement B*, p. 58]. The first number given is the sample size, the second, the number of defectives. (We shall not correct the sample size by subtracting the number of defectives, because in the one case where this might make a visible difference there are no defectives.) We divide the sample size by 10, and plot the triangles in Figure 4 (to this scale, the triangles reduce to vertical segments).

ADJUSTMENT IRREGULARITIES, ELECTRICAL APPARATUS

Lot	Sample Size	Defectives	Lot	Sample Size	Defectives
1	600	2	11	1550	7
2	1300	2	12	950	2
3	2000	1	13	950	5
4	2500	1	14	950	2
5	1550	5	15	35	0
6	2000	2	16	330	3
7	1550	0	17	200	0
8	780	3	18	600	4
9	260	0	19	1300	8
10	2000	15	20	780	4

From Figure 4 we measure the vertical deviations from the $\bar{p}$ line (which is assumed to be 0.27% based on past experience) and plot them on a regular control chart (Figure 6), being sure to keep the data in the order in which they originally appeared. (The use of tracing paper makes this process very easy.) In practice each new observation would first be plotted on binomial probability paper (perhaps at an enlarged scale) and then transferred. If the data are retained on the original probability paper, the advantage of examining the data for trends and runs would be lost.

Critique of Example 16. The control chart in Figure 6 looks very different from that usually given. In the usual chart Lot 19 is shown as beyond the control limits on the high side, and Lots 4 and 7 are not detected as being possibly too defective-free because we find that there can be no lower control limit. This kind of plotting might be useful even when the samples are of constant size.

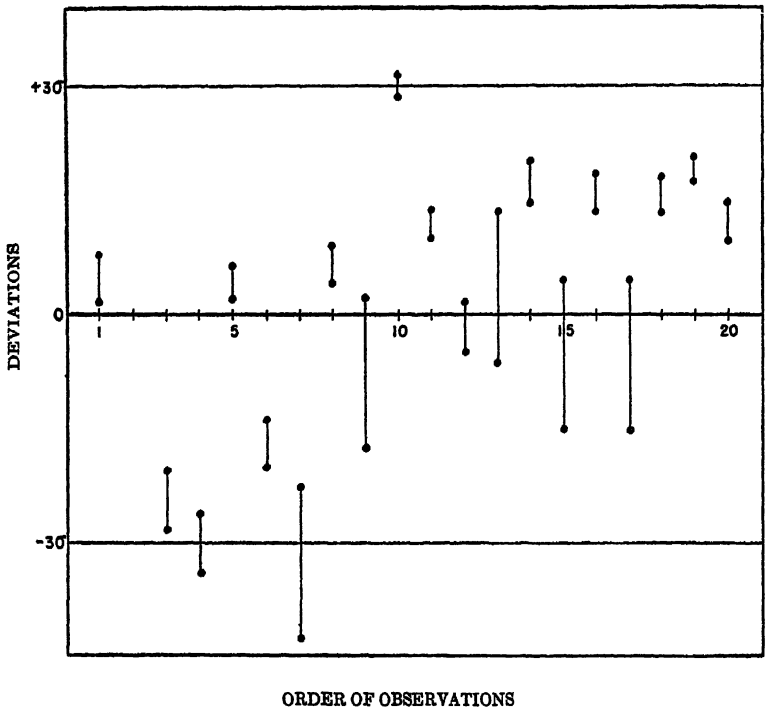
TESTS OF HOMOGENEITY

The general case

Given 5 or 50 or 500 paired counts which we wish to test for homogeneity, to test if it is reasonable that they have arisen from sampling

FIGURE 6

SHOWING COMPLETED, STABILIZED *P*-CHART AT ENLARGED SCALE (WITH SEGMENTS MADE VERTICAL).



a population with the same percentages in the two categories, the problem is the same, but the practical solution is different. In every case we plot the individual paired counts and draw the best fitting split, either by eye or through the sum of the paired counts.

We shall discuss three methods here, namely:

(1) graphical chi-square,
(2) range,
(3) counts in $\pm 1\sigma$ and $\pm 2\sigma$ strips.

Each of these has advantages and disadvantages, and, to the best of our knowledge, they can be compared as follows:

Method	Number of Samples		Advantages	Disadvantages
	Feasible	Recommended		
χ^2	2 to ∞	2 to 8 or 15	efficiency	labor
range	2 to ?	2 to 20	ease and speed	limited efficiency
counts	15 to ∞	15 or 20 to ∞	80% efficiency; relative simplicity	

The range is only recommended for 20 paired counts or less since its use for larger k involves the delicate details of the normal distribution and since its efficiency is less than the counting method.

To apply the χ^2-test, plot the paired counts and the split through their sum, and combine the *middle* distances by crab addition as explained at the end of Section 4.

Example 17 (See Figure 4). The following classic data by C. Goring quoted by K. Pearson and by M. G. Kendall compare the number of alcoholics and non-alcoholics among criminals according to crimes committed. The first number in each pair is the number of alcoholics:

Arson	(50, 43)	Stealing	(379, 300)
Rape	(88, 62)	Coining	(18, 14)
Violence	(155, 110)	Fraud	(63, 144)
		Totals	(753, 673)

The graphical display shows very clearly that (1) the observations are discrepant, (2) the crime of fraud is the only one for which the proportion of alcoholism is really different. Graphical chi-square computation gives 30.2 on 5 degrees of freedom—highly significant. When criminals convicted for fraud are removed from consideration the remaining five groups are each less than one standard deviation away from the new fitted line (690–539 split). Stealing is slightly misplotted.

Critique of Example 17. If the definition of "alcoholic" were sufficiently objective, and if the sample of convicted criminals represents a random sample of criminals, then the analysis seems sound. It cannot, of course, throw any appreciable light on the connection between alcoholism and crime in general, bearing only on the question "exclud-

ing fraud, do alcoholics tend to be convicted of some types of crime and non-alcoholics of others?"

A quicker method of analysis, and one well suited for drawing lines by eye is to compute the range of the sample, that is the sum of the greatest middle distances to the right and left of the line. Then the range measured in millimeters or standard deviations can be compared with Table 4 (at end of paper) to discover whether the samples deviate enough among themselves to provide evidence that the observations did not arise from random sampling from a single proportion in the population.

Example 18 (See Figure 4). In testing the effect of the X-chromosome inversion B^{M1} on secondary non-disjunction, K. W. Cooper (personal communication) raised 21 cultures of *Drosophila melanogaster*, crossing v-In(1)$B^{M1}/y^2w^av/Y$ females with wild-type Canton-S males. The presence or absence of secondary non-disjunction can be detected in female progeny. The 21 cultures gave the following results:

(9,135), (7,115), (11,118), (13,89), (15,148), (8,91), (6,113), (11,104), (9,122), (10,90), (15,155), (14,138), (5,84), (11,128), (2,34), (4,73), (4,107), (9,107), (10,103), (11,115), (8,104)

where the smaller class showed secondary non-disjunction. The split is drawn through the total of (192, 2273). The range is 16.5 mm which is far from significant—thus this test gives no evidence of heterogeneity.

Critique of Example 18. Clearly an amount of heterogeneity large enough to affect the estimated standard error of the grand mean would almost surely have been detected. The test seems adequate for its purpose.

A more refined, but still simple test is obtained by drawing parallel lines, at ± 5 mm and ± 10 mm. In sets of 21 or more samples, we expect about 5% outside the outer lines and about 33% outside the inner lines. The weighted sum

$$12 \text{ (no. outside 10 mm)} + 3 \text{ (no. between 5 mm and 10 mm)}$$
$$-2 \text{ (no. inside 5 mm)}$$

is distributed with mean nearly zero and variance nearly $11.72k$ in case of k homogeneous samples. Approximate significance levels are $5.65\sqrt{k}$ and $8\sqrt{k}$ for the 5% and 1% points.

Example 19 (See Figure 4). Returning to the data of Example 18 and drawing the 5 mm and 10 mm lines, we find (classifying borderline cases according to the center of the hypotenuse of the triangles).

	Expected	Found
Outside	1	0
Between	6	4
Inside	14	17

The weighted sum is 12–34 = −22 which is far from the significance levels of 25.9 and 36.5. This more delicate test finds no evidence of heterogeneity in the per cent of secondary non-disjunction in Cooper's cultures. Indeed, if anything the data are a little too homogeneous, though not enough to notice. These methods can also be applied to the case of unsymmetric counts, as in the following example.

Example 20 (See Figure 5). In testing the effects of X-chromosome inversions on primary non-disjunction, K. W. Cooper (personal communication) crossed 847 males with females of eight different chromosomal sequences. Exceptional cases can be detected in both males and females. The observed counts for males were (2885, 13), (7172, 18), (4672, 13), (9162, 14), (1389, 2), (2961, 4), (2199, 2), (1195, 1). Does the rate of primary non-disjunction seem to be constant?

The total count is (31635, 67), and the corresponding split together with the individual counts are plotted on Figure 5, where *all horizontal coordinates have been divided by 50.* The lines parallel to the total split are at ± 5 mm and ± 10 mm *vertically*. The range, measured *vertically* (since the horizontal coordinate has been reduced), is 19.3 mm which is not far from the 5% point of 21.8 mm.

Critique of Example 20. The method of Example 18b should not be applied on so few points, but if it were applied the weighted sum would be $12(1)+3(3)-2(4)=13$. The value of $5.65k$ is 16.0, but direct calculation of the 5% point yields 19. Thus the approximate 5% point is not too accurate for 8 points (16.0 is about the 10% point). Calculation of χ^2 by crab addition of the *vertical* deviations from the line to the total yields 13.9, which is again not quite at the 5% level for 7 degrees of freedom.

The four-fold table

The four-fold table, where a sample is classified into two categories in each of two ways has received very much attention by both applied and theoretical statisticians. Different methods of analysis have been given, some of which assume that

(1) the sample is a representative of samples in which only the total is fixed, or

(2) the sample is a representative of samples in which one set of marginal totals are fixed, or
(3) the sample is representative of samples in which all marginal totals are fixed.

Many of the "control group versus experimental group" experiments so common in biology, medicine, psychology, and education fall under (2), since the numbers in the control and experimental groups are fixed. Such experiments can be approximately analyzed as a homogeneity test as in the last section. For the case of *two* paired counts, the chi-square and range methods are equivalent, and the range is simpler.

Example 21 (See Figure 5). English et al. [7, 1940] took samples of 208 smokers and 208 non-smokers and investigated the incidence of coronary disease. They found (198,10) and (206, 2), where coronaries are the second category. The range is 17 mm which is significant at the 5% level.

They also took 187 cases with coronary disease, and 302 without, and investigated the incidence of smoking. They found (149,38) and (115,187), where smokers are the first category, which yields a range of 30.1 mm (Figure 5) which is horribly significant.

Critique of Example 21. These last two samples can be united into a four-fold table, but, in view of the way in which the data were obtained,

	Coronary disease	
	Yes	No
Smokers	149	187
Non-smokers	38	115

it would be *incorrect*, to compare (187,149) with (115,38) by this method and to assume that two binomials were being compared. However, the range obtained in this way is 29.7 mm and it is possible that such inverted tests on binomial probability paper give approximately correct answers.

A less obvious example

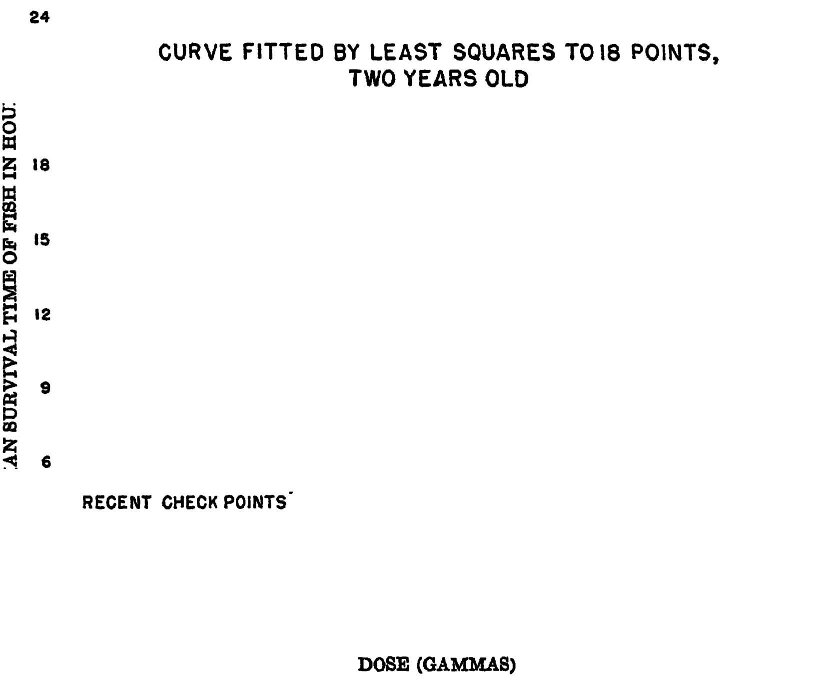
The ideas behind the sign test may be extended to give approximate tests in many situations of greater complexity. Such tests may be very useful, when used with the knowledge that they are quick, but often lack the sensitivity of more complex methods.

Example 22 (See Figures 7 and 5). A routine bioassay had been in

use for two years using a standard curve. Occasional checks on the standard had been made. The situation is shown in Figure 7, which raised two questions: (i) Does the curve agree with the recent points? (ii) If not, has something surely changed in two years, or may the difference be assigned to the combined sampling fluctuations in establishing and checking the standard curve?

FIGURE 7

BASIC BIOASSAY DATA FOR EXAMPLE 22 (A LESS OBVIOUS EXAMPLE).



The first question is answered by the split test, for comparing (19, 5) with a 50-50 split yields (Figure 5) a separation of 15 mm which is very highly significant.

The second question can be approximately answered as follows: the original least square fit to 18 points was probably more accurate than fitting a median to 18 points and less accurate than fitting a median to 36 points. A roughly fair test should come between a comparison of (19, 5) and (9, 9) and a comparison of (19, 5) with (18, 18). These give (Figure 5) ranges of 14.5 mm and 16.5 mm, which are both beyond the

5% level, indicating that the activity, or the fish, or the technique has probably changed.

Critique of Example 22. While these are not the most thorough tests which can be applied to this situation, anyone familiar with bioassay computation will appreciate their speed, simplicity, and clarity.

PART V—REFERENCE MATERIAL

MODIFICATION OF THE ANGULAR TRANSFORMATION

The original angular transformation

The angular transformation was introduced by R. A. Fisher in 1922 [9, p. 326] in a genetic situation where a certain proportion was varying by random fluctuation from generation to generation. In 1936, Bartlett [2, p. 74] proposed its use on experimental data as a means of stabilizing the variance when binomial data were subjected to the analysis of variance. Various authors have proposed its use for various purposes, a considerable number of references may be found in [13, 1947].

Bartlett's modification

In his 1936 paper, Bartlett also proposed an empirical modification to make the transformation more effective near $p=0$ and $p=1$. This was the device of transferring $\frac{1}{2}$ of a unit from the larger count to the smaller count. Thus (3, 29) would become (3.5, 28.5). This proved to be helpful, but had the annoying feature that both (3, 4) and (4, 3) were converted to (3.5, 3.5) which did not seem appropriate.

The smooth version

The smooth way of obtaining the good effects of $\pm\frac{1}{2}$ near the ends and ± 0 in the middle is to add $\frac{1}{2}$ to each cell, thus passing from $(n\text{-}k, k)$ to $(n\text{-}k+\frac{1}{2}, k+\frac{1}{2})$. It is clear that for values of p near 0, $\frac{1}{2}$, and 1 this will stabilize the variance very well. How well requires a numerical study, now in progress.

Correction for continuity

Most of the applications of binomial probability paper discussed above deal with tests of significance rather than with scoring paired counts. We must try, then, to assign nearly normal deviates, not to single paired counts, but to tails—to all $(n\text{-}k, k)$ for which $k \geq r$, for example. This is closely connected with the scoring problem, since a natural dividing line between $(n\text{-}r+1, r\text{-}1)$ and $(n\text{-}r, r)$ is $(n\text{-}r+\frac{1}{2},$

r-$\frac{1}{2}$), and in accordance with the last paragraph, this is to be scored as if it were (n-r+1, r). Thus we expect to find that

$$\sqrt{n+1}\left(\sin^{-1}\sqrt{\frac{r}{n+1}} - \sin^{-1}\sqrt{p}\right)$$

is nearly the normal deviate associated with the probability that $k \geqq r$, where k is binomially distributed according to n and p.

Flattening

Since the angles involved are rather small, it is plausible to replace them by their sines. This is of course what has been done in the examples, where we have always measured distances perpendicular to the splits. A little trigonometry shows that the distance from (n-r+1, r) to the p-q split is (in standard deviations)

$$2(\sqrt{p(n-r+1)} - \sqrt{qr}).$$

(To obtain distances in millimeters, replace 2 by 10.16 mm.)

Accuracy

The accuracy of the over-all approximation to

$$Pr\ \{k \geqq r \mid k \text{ binomial } (n, p)\}$$

by

$$Pr\ \{x \geqq 2(\sqrt{p(n-r+1)} - \sqrt{qr} \mid x \text{ unit normal}\}$$

has been studied numerically, and a note giving details will be submitted to the Annals of Mathematical Statistics (by Murray F. Freeman and John W. Tukey). The general conclusion is that the approximation is extraordinarily good near the 1% to 5% points, and remarkably good in general.

THE INCOMPLETE BETA AND F DISTRIBUTIONS

The binomial distribution is, as is well known, given by the expansion of

$$(q+p)^n \qquad q = 1 - p$$

where n is the number of cases, p the chance of a "success" and the term

$$\binom{n}{r}q^{n}$$

in the expansion is the probability of exactly r successes. The probability of r or more successes is given by

$$S = \sum_{x=r}^{n} \frac{n!}{x!(n-x)!} (1-p)^{n-x} p^x.$$

Using the well known device of differentiating both sides with respect to p and summing, we get

$$\frac{dS}{dp} = \frac{n!}{(r-1)!(n-r)!} p^{r-1}(1-p)^{n-r}.$$

Replacing p by t and integrating S from 0 to S, and t from 0 to p, we have the usual relation

$$\begin{aligned} Pr\ (r \text{ or more successes}) &= I_p(r, n-r+1) \\ &= \frac{n!}{(r-1)!(n-r)!} \int_0^p t^{r-1}(1-t)^{n-r} dt \end{aligned}$$

where $I_p(m, n)$ is the incomplete Beta-function. Hence if binomial probability paper successfully represents the binomial distribution it also successfully approximates the incomplete Beta-function.

Thus

$$\begin{aligned} &I_p(r, n-r+1) \\ &\qquad \sim Pr\ \left\{ x \geqq 2(\sqrt{p(n-r+1)} - \sqrt{qr}) \mid x \text{ unit normal} \right\} \end{aligned}$$

which seems, incidentally, to be a new analytic approximation to the incomplete Beta-function. Simplifying notation, we find that $I_p(m_1, m_2)$ corresponds to the distance from p-q split to the point (m_1, m_2).

The ratio of two independent mean squares obtained from normal variates of the same variance is Snedecor's F, which is related to Fisher's z by $F = e^{2z}$. The ratio of the numerator sum of squares to the total sum of squares may be written in terms of F as

$$x = \frac{n_1 F}{n_2 + n_1 F},$$

and its distribution is given by,

$$Pr\ \{x < p\} = I_p(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2).$$

Hence, to the approximation of binomial probability paper, a ratio

$$\frac{s_1}{s_1 + s_2} = p$$

of *sums of squares* has a probability of arising from populations of equal variance which is given by the tail area corresponding to the deviation of the count $(\frac{1}{2}n_1, \frac{1}{2}n_2)$ from the line $p-(1\text{-}p)$ which is the same as the line $s_1 - s_2$.

PART VI—INDEX, OUTLINE AND TABLES

Introduction

Table 1 is not intended to replace the worked examples, but rather to serve as a key for the new reader and a reminder for the old.

The short tables which follow are of standard distributions based on the normal distributions. Since millimeters are convenient units for use with binomial probability paper, they are given in both millimeters and in standard deviation units.

For maximum accuracy, use a sharp pencil! (Regular thickness automatic pencils may serve for some routine work, but finer lead will give better results.) The figures have been drawn for clarity, not accuracy.

Remember these methods are all approximations.

TABLE 1

INDEX AND OUTLINE

Example	Aim	Plotting Required	Remarks
	Part II. Plotting one observed quantity		
1 p. 182	Observed and theoretical proportions	1 paired count 1 split (theory)	Use short distance for significance level. Use both short and long for significance zone.
2 p. 184	Sign test	1 split (50-50) 1 paired count	
3 p. 186	Confidence limits for proportion	1 paired count 2 splits (at distance)	Use short distance
4 p. 186	Confidence limits for median	1 split (50-50) 2 paired counts (at distance)	Use short distance
5, 6 p. 187, 189	All F tests	1 split (sums of squares) 1 point ($\frac{1}{2}$ degrees of freedom)	
7 p. 189	Angular transformation	1 paired count 1 split (through middle)	

TABLE 1 (*Continued*)

Example	Aim	Plotting Required	Remarks
		Part III. Application to design	
8 p. 191	Designing binomial experiment	2 splits (theory) 2 parallel lines (at distance)	Distances correspond to one-sided significance levels at percentages to be controlled. (AQL and RQL=LTPD)
9 p. 192	Designing single sampling plan	2 splits (theory) 2 parallel lines (at distance)	
10 p. 192	Operating characteristic of sign test	1 paired count 1 split (at distance)	Use short distance
11 p. 193	Sample size for population tolerance limits	1 split 1 parallel line 1 horizontal line	Split-to-line distance desired confidence. Sum of counts-in determines horizontal
12 p. 194	Tolerance limits for second sample	1 split 2 paired counts (touching)	Split through 1st and 2nd sample sizes; distance to common vertex=confidence
13 p. 194	Operating characteristic of anova II	1 point (½ degrees of freedom) 2 splits (at distance) compute from ratio of split ratios	
		Part IV. Several paired counts	
14 p. 195	k proportions and a theoretical proportion	k paired counts 1 split (theory) 2 parallel lines (± 10 mm)	Expect 1 in 20 outside by middle distance.
15 p. 198	k proportions and a theoretical proportion	k paired counts 1 split (theory)	Combine middle distance by crab addition, (see p. 179)
16 p. 199	Stabilized p-chart	k paired counts split (assumed level)	Transfer to tracing paper as control chart
17 p. 201	Homogeneity of k proportions (k small)	k paired counts 1 split (sum)	Combine middle distances by crab addition (see p. 179)
18 p. 202	Homogeneity of k proportions (k 20)	k paired counts 1 split (sum)	Range of middle distances
19 p. 202	Homogeneity of k proportions (k large)	k paired counts 1 split (sum) 4 parallels (± 5 mm, ± 10 mm)	12 outside+3 between−2 inside 5%, 5.65 k; 1%, 8 k. Use middle distances.
20 p. 203	Homogeneity of k unsymmetrical proportions	as 17 or 19 with large count divided	As 17 or 19 with distances in undivided direction
21 p. 204	Four-fold table	2 paired counts 1 split (sum)	Range from middle distances

TABLE 2

MILLIMETER TABLE FOR NORMAL DEVIATE

Significance Level one-sided	Significance Level two-sided	Normal Deviate millimeters	Normal Deviate multiples of
50%	100%	0.0	.00
40%	80%	1.3	.25
30%	60%	2.7	.52
20%	40%	4.3	.84
16.5%	33.2%	5.0	.97
10%	20%	6.5	1.28
5%	10%	8.4	1.65
2.5%	5%	10.0	1.96
1%	2%	11.8	2.33
0.5%	1%	13.1	2.58
0.1%	0.2%	15.7	3.09
Conversion relation		5.080	1.000

TABLE 3

MILLIMETER TABLE FOR CHI-SQUARE

Degrees Freedom	Undoubled Millimeters At an Upper Significance Level of (50%)	5%	1%	Multiples of σ^2 At an Upper Significance Level of (50%)	5%	1%
1	(3.4)	10.0	13.1	(.5)	3.8	6.6
2	(6.0)	12.4	15.4	(1.4)	6.0	9.2
3	(7.8)	14.2	17.1	(2.4)	7.8	11.3
4	(9.3)	15.6	18.5	(3.4)	9.5	13.3
5	(10.6)	16.9	19.7	(4.4)	11.1	15.1
6	(11.8)	18.0	20.8	(5.3)	12.6	16.8
10	(15.5)	21.8	24.4	(9.3)	18.3	21.2
15	(19.3)	25.4	28.1	(14.3)	25.0	30.6
30	(27.5)	33.6	36.2	(29.3)	43.8	50.9

TABLE 4

MILLIMETER TABLE FOR NORMAL RANGES

Number of Observations	Millimeters At Upper Significance Level of (50%)	5%	1%	Multiples of σ At Upper Significance Level of (50%)	5%	1%
2	(4.8)	14.1	18.5	(0.95)	2.77	3.64
3	(8.1)	16.9	20.8	(1.59)	3.34	4.10
4	(10.0)	18.5	22.2	(1.97)	2.65	4.38
5	(11.5)	19.6	23.2	(2.26)	3.87	4.59
6	(12.6)	20.5	24.1	(2.47)	4.04	4.74
7	(13.5)	21.2	24.8	(2.65)	4.18	4.87
8	(14.2)	21.8	25.3	(2.79)	4.29	4.98
9	(14.8)	22.3	25.8	(2.91)	4.39	5.07
10	(15.4)	22.7	26.2	(3.02)	4.48	5.15
15	(17.4)	24.2	27.6	(3.42)	4.79	5.44
20	(18.8)	25.4	28.6	(3.69)	5.01	5.64

REFERENCES

[1] American Society for the Testing of Materials, "Manual on Presentation of Data," Philadelphia (1940).
[2] M. S. Bartlett, "The square root transformation in analysis of variance," *Suppl. J. Roy. Stat. Soc.*, Vol. 3, 1936, pp. 68–78.
[3] M. S. Bartlett, "The use of transformations," *Biometrics*, Vol. 3 (1947), pp. 39–52.
[4] W. J. Dixon and A. M. Mood, "The statistical sign test," *Journal of the American Statistical Association*, Vol. 41 (1946) pp. 557–566.
[5] Acheson J. Duncan, "Detection of non-random variation when size of sample varies," *Industrial Quality Control*, Vol. 4 (1947–48) No. 4, pp. 9–12.
[6] Churchill Eisenhart, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947) pp. 1–21.
[7] J. P. English, F. A. Willius, and J. Berkson, "Tobacco and coronary disease," *J. Am. Med. Assn.*, Vol. 115, pp. 1327–1328 (1940).
[8] C. O. Fawcett, et al., "A second study of the variation and correlation of the human skull, with special reference to the Naqada Crania," *Biometrika*, Vol. 1 (1902) pp. 408–467.
[9] R. A. Fisher, "On the dominance ratio," *Proc. Roy. Soc. Edinburgh*, Vol. 42 (1922) pp. 321–341.
[10] R. A. Fisher and K. Mather, "The inheritance of style length in *Lythrum salicaria*," *Annals of Eugenics*, Vol. 12 (1943) pp. 1–23.
[11] R. A. Fisher and K. Mather, "A linkage test with mice," *Annals of Eugenics*, Vol. 7 (1936) pp. 265–280.
[12] Frederick Mosteller and John W. Tukey (designers), *Binomial Probability Paper*, Codex Book Company, Norwood, Mass. 1946.
[13] Statistical Research Group, Columbia University, *Selected technique of statistical analysis*, McGraw-Hill, 1947.

TEACHING STATISTICAL QUALITY CONTROL FOR TOWN AND GOWN*

EDWIN G. OLDS
Carnegie Institute of Technology
AND
LLOYD A. KNOWLER
State University of Iowa

During the next five years, in manufacturing plants and in the engineering schools, there will be many new programs initiated to meet the demands for education in Quality Control by Statistical Methods. This paper has been written with the hope that it will be of some help in the planning and executing of such programs. Beginning with a description of types of courses and a discussion of possible content, it touches on the questions of who is to do the teaching and how the subject matter might be presented and motivated. The importance of follow-up work is stressed. The latter part of the paper discusses the choice and utilization of instructional aids and describes some of the materials now available.

I—TYPES OF COURSES

THESE REMARKS concerning the teaching of Statistical Quality Control will be confined to four general types of courses, each of which permits various subdivisions. The four general types will be referred to as: (1) intensive or so-called ten-day courses; (2) extension and evening courses which, by nature, usually are given on a part-time basis; (3) university or college credit courses; and (4) in-plant training courses. Naturally these types are not necessarily mutually exclusive.

Intensive Ten-Day Courses.

Although statistical work with special reference to applications in industry has been given in colleges and universities for many years, it is believed that it received a great impetus during World War II because of the thirty-odd so-called eight-day courses in Quality Control by Statistical Methods most of which were sponsored, to a large extent, by the War Production Board and the United States Office of Education under the Engineering, Science, and Management War Training Program in cooperation with various educational institutions. Repre-

* Presented at Section on Training of Statisticians of the American Statistical Association joint with American Society for Quality Control and Institute of Mathematical Statistics at Cleveland, Ohio, on December 27, 1948. Parts I and II were presented by Lloyd A. Knowler, Part III, under a separate title, was presented by Edwin G. Olds.

sentatives from government and industries engaged in production directly related to the war effort were permitted to attend these tuition-free courses. Many of the trainees in the short courses received, for the first time, any formal background in statistical procedures. Because of some continued need for the type of training which can be given in these courses, a few educational institutions have continued to sponsor similar courses on a moderate tuition or fee basis.

The basic course has generally been expanded to ten days. A typical ten-day intensive course has the first day set aside for executives. The Executives' Session is devoted to an explanation of the aims and possibilities of a quality control program and procedures for the installation of such a program in the plants is indicated. In addition to the executives the session is also attended by the trainees who expect to remain the entire ten days.

The program during the remaining nine days consists of a series of conferences, lectures, and laboratory periods. In addition to daily sessions, three or four evening sessions are scheduled. Instructors in the courses are continually bombarded with questions during the short intermissions between sessions. General lectures are given to the trainees. The discussion of specific problems and the working of laboratory exercises is facilitated by dividing the general group into small sections. While the emphasis is to apply the statistical method to design, specifications, production, and inspection, the evening sessions and informal discussions during the day are often devoted to additional phases of statistical work.

The course content in the intensive courses is ordinarily divided into two parts: (1) control charts; and (2) acceptance sampling.

The use of control charts of the usual types ($\overline{X}$, R; p; np; c; and $\bar{c}$) as a production tool is stressed; in fact, the comparison of a control chart to a highway, the descriptions of a control chart as a picture, a newsreel, or as an advertisement of the workers' product is quite common.

In studying a control chart it is usually necessary to introduce the concept of a frequency distribution. Following this, or in connection therewith, a control chart is constructed based upon shop data. This is followed by a laboratory exercise and discussion. Thereafter demonstrations of the effect of changes in average or range are given with the use of chips or beads as indicated in the second phase of this report. This general pattern of instruction is altered occasionally to fit the particular needs of the trainees in the course.

The facts that new uses are being discovered constantly and that

additional concerns are making use of statistical procedures as more individuals are being trained are emphasized. It is pointed out that a single chart on a single machine for a single operator in a large corporation is essentially a one man company. The value of a control chart for short runs and its value when applied to new lines of production is more than noted. Quality and economy are stressed. The diagnostic value and the predictive value of determining when and where to look for trouble are also stressed.

The introduction of acceptance sampling in the basic course pertains largely to the Army Ordnance tables, the Dodge-Romig tables, and an introduction to sequential analysis together with a demonstration and discussion of the defects of 10 per cent sampling.

Some attention is given to the construction of a sampling plan, largely to bring out the value of an operating characteristic curve and an average outgoing quality curve. The use of a certified control chart in the manufacture of goods to replace acceptance sampling, or as another type of acceptance sampling, has been advocated by the presidents of at least two large corporations in this country who use and who realize the value of control charts.

In addition to a regular staff in charge of the course, three or four periods are given over to representatives from different industries using statistical quality control who discuss the practical applications in their own control problems.

It has been found fruitful, in some instances, to maintain a rather aggressive follow-up program including at least two two-day clinics in which the trainees attending the course meet with representatives of the instructional staff to discuss the application of quality control to specific manufacturing problems, to check on the correct application of statistical procedures, and to provide an opportunity for exchanging information. These follow-up meetings are exceptionally well attended. In some instances additional statistical topics such as correlation, analysis of variance, design of experiments, chi-square, and other tests of significance are considered. For the greater part, these additional techniques are taken up in section meetings which tend to form as a result of the short courses.

It has also been found desirable to give so-called advanced courses of eight days' duration. The topics included in a typical course are: a review of control charts; significance of differences; analysis of variance; correlation—linear and multiple; further aspects of acceptance sampling; further aspects of sequential analysis; chi-square; and use of calculating machines. An instructional pattern, similar to that used for

the elementary courses, is followed and also follow-up programs are maintained.

The usual experience in these intensive courses is to find that, initially, representatives from industries want and demand only so-called practical materials. After taking the course and making some uses of their procedures, they learn the value of statistics and then decide that the difference between practice and theory is rather fuzzy, if a difference even exists. In fact, many take the view that what works in practice does so because it has a reason and that the reason is the theory. It is these "practical" persons who ordinarily request additional "theoretical" work. In fact, they often ask that what might be a semester's or year's work in mathematical statistics be presented during one forenoon, possibly in one hour; not only be presented but be set out in such a way that they can go back to their plant and use it in the particular problem at hand.

Extension and Evening Courses.

Many of the extension and evening courses given once or twice a week, occasionally once a month, follow the same general pattern as the intensive courses although to some extent many of the discussions are a little more formalized. Naturally the material on each topic tends to be more self-contained to the extent that it is usually unnecessary to attend the first lecture or two in order to take something away from the later ones.

It is encouraging to note that enrollment tends to be maintained in these evening courses even though a year may be spent in one series. It is also encouraging to note that the same individuals find it profitable to repeat a series so as to get new ideas from different lecturers or instructors.

There is a trend toward offering an advanced, as well as an elementary, series of lectures.

University and College Credit Courses.

The content of university credit courses may not be so clearly defined. Because of the broad aspects of mathematical statistics, each person who is instructing a course of this type has had different experiences; it would be natural for them to draw upon their experiences to a considerable extent. It is believed, however, that through discussions such as this an approach to standardization will be effected.

The catalog description of such a course might read: "The theory and applications of that part of mathematical statistics used in main-

taining control of the quality of a manufactured product or of a service; in the construction and use of acceptance tests, and the associated concepts of the operating characteristic curve and the average outgoing quality"; or references might be made to "Elementary statistical methods and their application to industrial problems; construction and interpretation of Shewhart control charts; Sealy's modified techniques; Dodge-Romig and Army Ordnance Tables for acceptance sampling; quality assurance for sampling by measurement; introduction to sequential analysis; methods of correlation; elementary analysis of variance."

A course in Statistical Quality Control should be given in that college or division where it is accessible to the student—not merely possible for him to register. It is relatively easy to effect this result in some schools—in others a special problem may be created. In so far as engineering colleges are concerned, a course in Statistical Quality Control might be given during the junior, senior or graduate year. In such instances the student will already have studied through the calculus, hence many more of the problems can be approached directly. The student, through his shop experiences, has a better idea as to the working of a machine, if not the working of industry.

For that reason, a three semester hour course following the general pattern of the recent avalanche of books on Statistical Quality Control, with such supplementary reports as the instructor desires, may orient the average student about as far as is necessary. The developments and illustrations of various principles would naturally be supplemented by demonstrations such as will be pointed out in the third part of this report.

Many students will observe the desirability of more work. Additional work in mathematical statistics may be taken immediately or after a year or two of experience. The additional topics which could be covered in a subsequent course or courses will be of considerable value for the quality control engineer of a large company. In fact, it is very likely that it will be of value to the quality control engineer in a smaller concern. It is not unusual for executive officers of large corporations to seek a quality control engineer with a background in shop and only a little work in statistics. After some understanding, however, the reverse is true. The demand for well trained statisticians is well known today. Much of this demand has been augmented, however, by initial applications under the direction of open-minded industrial executives. In fact, it might be observed that these persons have been most effective promoters and have also made possible many interesting prob-

lems for investigation. They have not been backward in advocating the need in their company for a person with training equivalent to a Ph.D. in mathematical statistics. These executives are particularly insistent, however, that the quality control engineer with such advanced training spend considerable time in shops and in assisting in the applications. They have realized that the main reason many executives and supervisors do not use mathematical statistics is that they have little or no knowledge of it.

In-Plant Training Program.

The in-plant training program is a very important phase of statistical quality control. Some of the larger companies have held a two-day training program for their top executives in which they attempt to give a rather broad overall picture of the subject. A program to achieve such an objective might be as follows: an introduction to statistical quality control; construction and interpretation of $\overline{X}$, R control charts; discussion of additional types of control charts; introduction to acceptance sampling; and suggestions on putting quality control to work. Executives are very much interested in such a program. They are particularly interested in the demonstrations which seem to have some relation to their plant operations.

Following this two-day program, the executives then select an individual from each plant in their company to take a ten-day course either sponsored by the company or by an educational institution. If within the company, the material considered is that pertinent to their own problems. The quality control engineers then return to their separate plants and make a pilot run as an application. They then train persons of supervisory and operator level in those particular aspects in which those persons happen to be most interested and for which they feel the greatest need. This may be the use of control charts for production or for design, or it may be some aspect of acceptance sampling. In the meantime, the quality control engineers join various statistical societies, start a library, and continue their study. It is not unusual for them to seek personnel from the educational institutions to assist them.

The procedure of having the in-plant training program divided into several periods wherein classroom discussions of a week or two are followed by a week or two of work in the plant, and the process repeated, has a great deal of promise.

II—PREPARATION OF TEACHERS

Since statistical quality control includes so many other modern statistical techniques in addition to control charts and acceptance sampling, it might not be out of order, in a paper such as this, to make a few remarks concerning the preparation of teachers. It is recognized that much depends upon the individual in charge of the course. In the first place, it is essential that the teacher have a knowledge of the subject and be interested in transmitting that knowledge or in getting another individual to gain similar knowledge. As with so many professions, the work of a quality control engineer is getting beyond the cook book style. It is rapidly becoming essential that the prospective teacher study two or three years of mathematical statistics, with one or two of these years in the Graduate College. In order for him to do this with much facility it is desirable, if not necessary, that he have a rather thorough background in pure mathematics. Also, the prospective teacher needs a background in some area such as engineering, business or industry from which to draw information. This is a large order, to be sure, but we need to realize that teaching is a profession and that it should be so recognized, both in the preparation required and in monetary rewards.

III—INSTRUCTIONAL AIDS

In the previous sections we have discussed what material might be taught, how it might be organized and motivated, and who should do the teaching. In this section attention will be directed toward the working tools needed and how they can be used to clarify quality control principles and develop power in analyzing manufacturing problems. Teachers in many fields seem to be capable of presenting their subjects acceptably when equipped with nothing more than a basic text, a portable blackboard, a stick of chalk and an eraser. For many, in fact, this seems to be the maximum list of requirements. Such does not seem to be the case for statistical quality control where successful teachers find effective use for a large collection of various kinds of paraphernalia.

Classroom and laboratory equipment can be classified under five general headings: (1) Textbooks; (2) Material for supplementary reading; (3) Problems; (4) Calculating machines; and (5) Gadgets. Some brief comments on each may prove helpful to embryo teachers.

Texts.

To remark that the success of a course depends, to a considerable extent, on the choice of text is distressingly trite, but in this case, it

seems necessary. When the Office of Production Research and Development initiated its statistical Quality Control Program early in 1943, pitifully few books on the subject were available, and most of them would be classified as treatises, rather than texts. Since the war, publishers have been able to put on the market a considerable number of books with the Quality Control label, but it seems fair to assume that most of the books were not prepared for use as basic texts in the field. Before any of them are adopted for the classroom the following questions should be considered:

1. Does the book give a sound exposition of the general philosophy and statistical principles basic to statistical quality control?
2. Is the exposition of sufficient breadth and completeness to meet the needs of the course?
3. Is the level of sophistication appropriate, in view of the students for whom the course is planned?
4. Will the book generate enthusiasm for statistical quality control as an engineering tool?
5. Is the arrangement and problem content such that effective day-by-day assignments can be made?

It would be surprising to find many books meeting all five tests satisfactorily. A book written to sell Quality Control to the busy executive would miss its aim if it attempted more than a brief overview of the simplest methods together with a briefer explanation of why they work. A manual prepared for use in the shop seldom will contain any derivation of the formulas being applied or hint as to many of their limitations.On the other hand, a book directed at advanced undergraduate or graduate engineering students with calculus, and perhaps some elementary statistics "under their belts" would be tragic for an in-plant training course at the supervisory level.

The rapidity with which manufacturing organizations have adopted statistical methods for the improvement of their operations can be attributed, to a large extent, to the missionary efforts of industrial men who were trained in short wartime courses. Reports of savings in manpower and materials, together with improvement in quality spurred them to study the methods, to apply them to their own problems, and to spread the gospel. There is no diminution of need to whip up enthusiasm but, with the frenzied haste of war work at an end, there is more time for careful appraisal of success stories. Furthermore, the serious student is more likely to be impressed by the powerful nature of statistical methods than by dollars saved. In no sense does this

comment imply that case studies should be deleted from a text or course, but rather that they should be chosen cautiously and imbedded in enough detail so that the student has some opportunity to check the correctness of the solution.

Students differ in the amount of verbiage they need, but almost without exception, they need problems to clarify and emphasize principles. Problems daily, like apples, seem to have a most beneficial effect. Texts without problems, or with problems involving excessive amounts of mechanical manipulation and calculation, place a heavy burden on student and teacher. It does not follow that, simply because a set of data originated in a plant, it will be an efficient hammer to drive home a principle.

One further remark on the choice of a text. A successful book or course does more than arouse interest, inculcate principles, and develop skills; it must create and foster the urge for continuation of the learning process; and it must chart the path and locate some of the important land marks. In other words the book should contain a generous and appropriate list of references which will carry the student beyond its boundaries in all directions.

Material for Supplementary Reading.

Even if a text with a satisfactory set of references is located, it is too much to expect either that all useful material will be included or that all of the papers listed will be readily available. For those readers who have examined university, city or company libraries this lack of availability need not be elaborated. Therefore, the teacher of statistical quality control will find it necessary to compile his own list for outside reading and find a way of giving his students easy access to any unpublished reports which they should see. The authors of this paper are much relieved that there is no room in this paper to include recommendations regarding books to purchase. Fortunately, many of the leading statistical, mathematical and engineering journals not only list new books received, but also print authoritative and prompt reviews of many of them. At least two bibliographies on statistical quality control [1] [2] have been compiled rather recently. Both books and periodicals are listed, but the field of coverage is largely restricted to control charting and acceptance sampling. Furthermore, both will be somewhat out of date by the time this paper reaches publication.

The bi-monthly journal, *Industrial Quality Control*, is the official publication of the American Society for Quality Control and, naturally, caters to the needs of its members. These needs are not greatly differ-

ent from those of beginning students in the field and, therefore, the publication is a rich source of material. The *Supplement to the Journal of the Royal Statistical Society* also specializes in industrial applications of statistics, but at a somewhat higher level. Also, useful papers will be found in the *Journal of the American Statistical Association* and *Biometrika*.

The *Annals of Mathematical Statistics* is devoted, primarily, to basic theory, but some of the theory is directly applicable to quality control problems, even under the narrowest meaning of the appellation.

A set of twelve *Quality Control Reports*, prepared at the close of the war under the auspices of the Quality Control Program at Carnegie Institute of Technology, is still available and can be obtained at a cost of two dollars from the Office of Technical Services, Department of Commerce, Washington. In the main, these are case histories of initiation of quality control methods.

Excellent papers on statistical methods and their applications appear from time to time in many scientific, engineering and trade journals. These may pass unnoticed unless students are briefed to watch for and report them. Several of the societies hold occasional symposia on statistical methods. Usually the reports of these meetings are worth locating. Unpublished plant reports are useful, particularly if prepared by former students.

Sources of Problem Material.

There is a dearth of satisfactory problem material for easy reference. Most teachers of statistical quality control do not hesitate to beg, borrow, or even purloin all of the industrial data they can. But most files contain pitifully few really good problems. Perhaps "good problem" should be better described. Mr. Wyatt Lewis, of the Ontario, California plant of General Electric Company, contributed such a problem for inclusion in the *Outline Manual for Quality Control by Statistical Methods*, which was written by Working and Olds and used by them and by many others in the teaching of intensive eight-day courses during the war. (The *Manual* has been available for two dollars from the same source as the *Quality Control Reports* mentioned above.) The problem is called the Rheostat Knob problem and is an application of the Shewhart Control Chart techniques. Data on an easily described quality characteristic is given for several days' production for which both specifications and manufacturing conditions were changed. Interesting details are supplied. The problem divides naturally into several parts, each of which can be assigned as a separate

exercise. When a calculating machine is used or when the numbers are coded, the time for computation is not excessive.

The above-mentioned manual has a few other good problems. Manuals prepared for intensive courses given more recently at various universities have others. Recent books have some. Others can be manufactured by judicious use of the data and circumstances given in technical articles. Good problems *can* be dug up, but a modicum of work is required—and, perhaps, a little harmless chicanery.

Calculating Machines.

Recently one of the authors read about an elementary course in statistics which was said to have been given quite successfully ("not ideally!") without the use of a computing laboratory. A similar remark might be made regarding beginning courses in statistical quality control. Obviously it is debatable and would not apply to courses at an advanced level. Statistical quality control *can* be done without calculating machines, and there seem to be two principal arguments for such a procedure: (1) economy; and (2) the distressing fact that some industrial organizations do not furnish machines for their quality control departments and so their men need to learn how to get along without them.

For the most effective teaching of statistical quality control the authors agree in favoring a computing laboratory with sufficient equipment to accommodate a class of students. Not all classes would be held in the laboratory but, when used for a class, each member would have an automatic calculator of the same model. Needless to say, the laboratory should be kept open at stated hours for individual work, with an assistant in charge, competent to teach the operation of the machines and responsible for their ordinary care.

It is useful to have late models of the principal kinds of calculators available, so as to broaden the students' experience. Punch-card equipment is a welcome addition. Several of the gadgets to be described in the next section are a proper part of laboratory equipment. So are statistical tables and some reference books. If space and supervision is adequate, there is little danger of collecting too much equipment.

Gadgets.

In days of fairly recent yore, the average teacher of elementary probability reached the heights when he had each of his students toss a coin 100 times, recording the succession of heads and tails. Then he was content to retreat to a consideration of either the few reports of

such experiments in the literature or "cooked-up" examples of the same type. It is hard to understand why he failed to appreciate the pedagogical value of designing an experiment to illustrate a point of theory, predicting the result, running the experiment, and then taking the consequences if it turned out wrong. Whatever the reason, it is fortunate for the field of statistical quality control, that its leading teachers have broken away from tradition and have shown no hesitation in using any available gadget which promised to assist in fixing a concept in the student's mind.

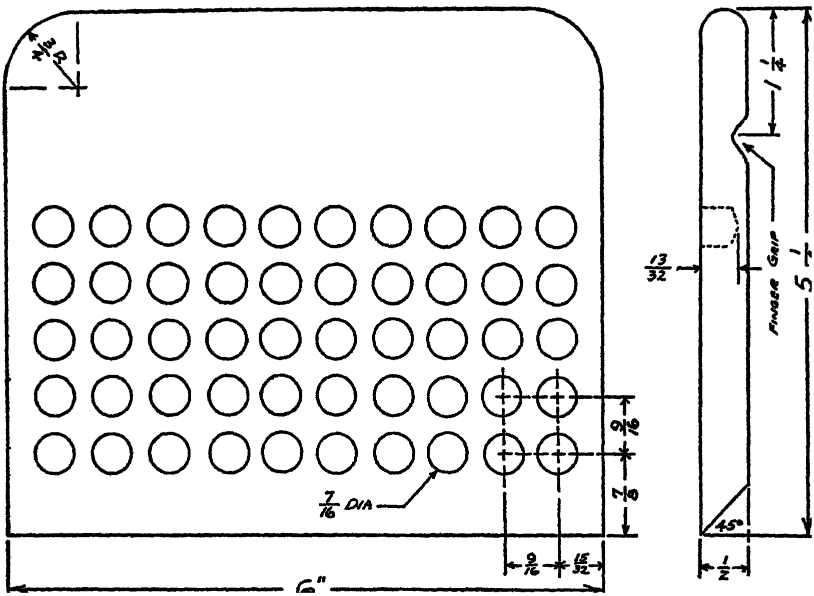


EXHIBIT 1
PLAN FOR 50 HOLE SAMPLING PADDLE
(adapted from a drawing by H. B. Rogers)

At some time during the last five or six years not only thousands of students, but scores of corporation presidents and their associates have been introduced to the vagaries of acceptance sampling by means of a box of beads and a 50-hole sampling paddle. The authors do not know who originated the paddle, but the man who made it famous was Holbrook Working. His paddle (see Exhibit 1) started as a wooden board about 6″×6″×½″. In it were sunk five rows of ten holes each, about 9/16″ apart, center to center, 7/16″ in diameter and 13/32″

deep. Near one edge the board was grooved to provide a handle. The rest of the equipment consisted of a cardboard box of several hundred 10-mm wooden beads (about 1200 white and 200 red). When Dr. Working went to the Walco Bead Company, 37 W. 37th St., New York City to purchase the beads and explained his purpose he met with astonishment, if not incredulity.

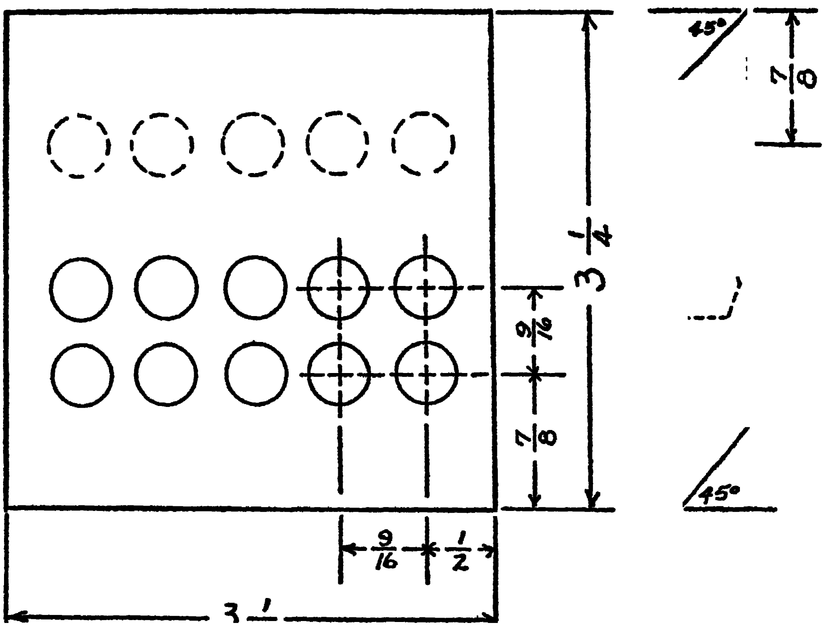


EXHIBIT 2
PLAN FOR 5 AND 10 HOLE SAMPLING PADDLE
(adapted from a drawing by H. B. Rogers)

The best known demonstration is concerned with illustrating the weaknesses of 10 per cent sampling. In the box 1152 white beads (good items) are mixed with 48 red beads (defective items). A paddle of beads (representing a lot from a controlled process) is withdrawn, the number of red beads counted, and the fifty beads dumped into a second box. Then, with a 5-hole paddle (see Exhibit 2), five beads are scooped from the fifty and number of defectives noted. The agreement is to accept the lot only if there are no defectives in the sample. Of course some of the lots accepted are worse than many of the lots rejected. Also the per cent defective in the uninspected portions of rejected lots

EXHIBIT 3

RECOMMENDED FREQUENCY DISTRIBUTIONS FOR SETS OF DISKS FOR SAMPLING FOR VARIABLES*

Set No.	1A	1B	1C	2	3A	3B	4
Color of Figures	Green	Black	Black	Blue	Black	Red	Red
Mean	0	0	0	+2	0	+1	+4
Standard Deviation	1.715	1.715	1.715	1.715	3.470	3.470	1.715
Numbers	Frequency						
−10					1		
− 9					1	1	
− 8					1	1	
					3	1	
					5	3	
					8	5	
− 4	3	3	3		12	8	
− 3	10	10	10	1	16	12	
− 2	23	23	23	3	20	16	
− 1	39	39	39	10	22	20	1
0	48	48	48	23	23	22	3
1	39	39	39	39	22	23	10
2	23	23	23	48	20	22	23
3	10	10	10	39	16	20	39
4	3	3	3	23	12	16	48
5	1	1	1	10	8	12	39
6				3	5	8	23
7				1	3	5	10
8					1	3	3
9					1	1	1
10					1	1	
11						1	
	200	200	200	200	201	201	200

* It is necessary to have sets designated as 1A, 2, 3A and 3B, or sets differing little from these as regards means and standard deviations Sets 1A, 2, and 4 will be thrown together for one demonstration, and bear figures in different colors to permit subsequent sorting. Sets 1B and 1C are used in a demonstration that calls for drawing from 5 bowls, 3 of which are alike.

The chips are white plastic discs, 9/16″ in diameter and 3/32″ thick. They can be purchased from Lamb Seal and Stencil Company, 824—13th Street, N.W., Washington, D. C. The wooden beads used in the sampling demonstrations are 10 mm. in diameter. They can be purchased from the Walco Bead Company, 37 W. 37th Street, New York City. A supply of 1200 white, 200 red, and 100 blue beads is suggested.

is about the same as the per cent defective in the uninspected portions of accepted lots, etc., etc.

It requires considerable restraint on the part of the authors to prevent them from devoting an entire paper to a description of the various uses of the paddle in connection with standard sampling tables, control

charts, tests of significance, confidence intervals and the like. A second paper could be written on "chips drawn from a bowl." White plastic chips about $\frac{5}{8}''$ in diameter and 3/16" thick are marked with numbers to simulate a normal distribution. Exhibit 3 gives suggested distributions for the various bowls which were found useful in O.P.R.D. courses. (These distributions seem to have been devised by Holbrook Working.)

Drawing samples of five from one of the bowls provides the data for a control chart. Changing bowls gives a graphic picture of the effect of a changed mean or standard deviation. With two bowls the meaning of Fisher's *t*-test becomes more clear. Correlate pairs of units and the presence of sample correlation when the universe correlation is zero is demonstrated in a fashion more telling than any appeal to theory. Use three bowls or more and analysis of variance can be introduced successfully. Possible demonstrations for stratified or sequential sampling are easy to plan and execute. The list of other useful demonstrations is almost unlimited.

Is it true that in random assembly, the square of the natural tolerance of an addition of several components is equal to the sum of the squares of the natural tolerances of the components? Professor MacCrehan has a set of 100 blocks to demonstrate that this is good theory. There are five sets of painted blocks; red, black, white, yellow, and green. In each set the twenty blocks vary in thickness. Assemblies are made against an upright board, on which are painted pairs of lines bounding the natural and absolute tolerances for the assembly. When five blocks, taken at random, one from each set, are piled against the board, the top of the pile falls between the two natural tolerance limits. Yet five blocks can be found which will reach the upper absolute tolerance line or just match the lower absolute tolerance.

A less spectacular, but none-the-less, useful gadget is a table of random sampling numbers. Several such tables are available [3], [4], [5]. While the same jobs can be done with a table as with the ships or beads, the use of a random sampling table at the elementary level does not seem to be very effective. At a higher level such a table is almost invaluable in connection with distribution theory. Suppose, for example, one one is forced to investigate the behavior of statistics in random samples from a very odd universe, one for which the density function might even be unknown. Numbers from the table can be so assigned that the universe is simulated in a form convenient for sampling. Then a large number of samples can be drawn easily and quickly, and the statistic under scrutiny can be calculated, tabulated, and studied. No further

work may be needed to arrive at satisfactory answers to many practical problems.

Having emphasized the point that gadgets are useful and lacking space for detailed description, the authors close this section with a list of several other physical aids to instruction. These gadgets, as well as those listed above can be used by students as well as teachers. In fact, students can learn some statistics by devising their own gadgets.

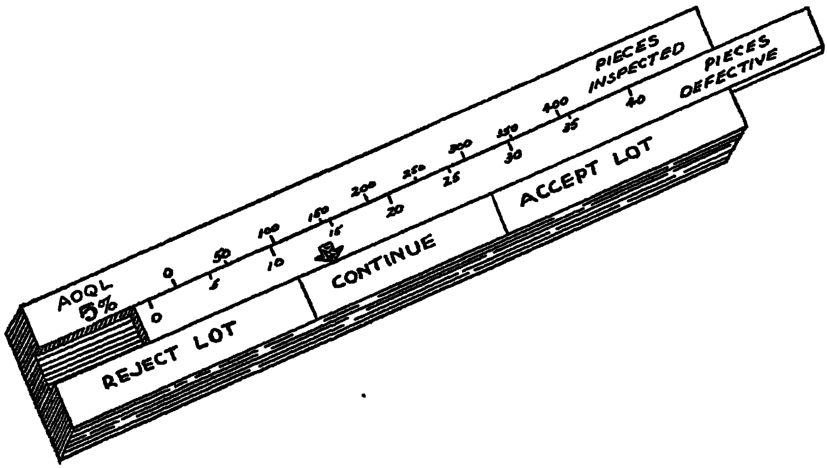


EXHIBIT 4

SKETCH OF A SLIDE RULE FOR SEQUENTIAL SAMPLING USED BY SERVEL, INC.
(from a photograph supplied by W. E. Gibbons)

1. Galton Board or Quincunx
2. Slot machines, seized in police raids
3. Sets of dice—fair and biased
4. Slides and movies
5. Colored plastic balls or ball bearings or marbles (These may be used with a paddle instead of beads. Both bearings and marbles are heavy and noisy.)
6. A can, for mixing chips, mounted on a phonograph turn table. The can is tilted so that the audience can observe the mixing procedure.
7. Pieces of wooden doweling of an assortment of lengths and diameters, with weights of pieces marked on them. This gadget is useful in explaining multiple correlation.
8. A sampling machine designed by T. H. Brown, and D. H. Leavens. Beads are mixed by revolving a closed container shaped like an

oil can. A sample is obtained by tilting the can so that beads roll into its transparent spout.

In addition to the gadgets built to provide for striking demonstrations of principles, the various mechanical devices concocted to help with practical application are deserving of mention. Exhibit 4 is a sketch of a slide rule being used effectively for sequential sampling. A slide-disk calculator has been designed for calculating standard deviations. Another type can be used in the shop to get control limits for average, range, and fraction defective charts. At least one quality control engineer inserted baffles in sheet metal cans of various sizes in such a fashion that they would mix lots of small parts and separate out random samples of specified sizes.

The authors have made no attempt to give an exhaustive list of gadgets. No doubt many readers have others of comparable merit. If so, it is to be hoped that careful descriptions of such gadgets, together with reasons for their construction and examples of their use, will be written and submitted for publication.

ADDENDUM

At the suggestion of a referee, who was present at the meeting when the above paper was presented and discussed, some brief comments are being added. They may delineate more clearly the nature of statistical quality control and indicate some of the main objectives of elementary courses in the field.

Statistical quality control is the application of statistical methods to the improvement of the manufacturing operation. At each stage in the life cycle of a product, from recognition of consumer need, through design, specification, fabrication and inspection, to final assurance of consumer satisfaction, statistical methods can play an effective role. All brands of quality control are concerned with the same problems but only *statistical* quality control utilizes *statistical* methods in solving them.

Industrial statistics and statistical quality control have many concepts and techniques in common but the dual classification is necessary. Many executives in charge of manufacturing regard industrial statistics as closely akin to business statistics and, therefore, mainly preoccupied with questions arising in sales and accounting. While these executives recognize the importance of such questions, they are primarily interested in problems of manufacturing. Therefore, they welcome any potential aid to quality control engineering which statistical methods may have to offer.

In a one-semester or two-semester course, or in intensive courses, there is neither the expectation nor the implication of transforming engineers into statisticians. The average engineer does not have the prerequisite mathematical background for much theoretical statistics. More preparation would be useful but it is a choice between taking him as he is or missing the opportunity to provide him with a few of the fundamental concepts and methods which he can grasp readily and apply with confidence. Most teachers try to pack as much statistics as possible into the time allotted. While statistical quality control should certainly reach beyond control charts and acceptance sampling by attributes, these two topics crowd a one-semester course. Parenthetically, this may be the reason why some engineers view statistical quality control as comprising only control charts and acceptance sampling.

In conclusion, it might be noted that there is little danger that a good teacher will encourage his students to view themselves as statistical experts. It seems to be more general to find students of statistical quality control very conscious of their limitations. This has two good effects: they are willing to seek advice and they are eager to learn more statistics.

REFERENCES

[1] "*Bibliography on Statistical Quality Control*," Report No. 11 of a series of Quality Control Reports published under the OPRD Quality Control Program, Carnegie Institute of Technology, Pittsburgh, 1945.

[2] G. I. Butterbaugh, "*A Bibliography of Statistical Quality Control*," University of Washington Press, Seattle, 1946.

[3] L. H. C. Tippett, "*Random Sampling Numbers*," Tracts for Computers, No. 15, Cambridge University Press, London, 1927.

[4] M. G. Kendall and B. Babington Smith, "*Tables of Random Sampling Numbers*," Tracts for Computers, No. 24, Cambridge University Press, London, 1939.

[5] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, Oliver and Boyd, London, 1938. Table 33 is a table of random numbers.

THE USE OF SAMPLING IN GREAT BRITAIN*

C. A. MOSER

Assistant Lecturer in Statistics, London School of Economics

The object of this article is to survey the use of sampling methods in Great Britain. The most important sample surveys undertaken by government departments, research organisations and commercial agencies are described with particular reference to their aims and sampling methods. Deficiencies in British sampling practice are discussed and suggestions for possible future developments are made.

INTRODUCTION

ONE OF THE most notable developments in the field of statistics in recent years has been the increasing utilisation of sampling methods in the study of human populations. Although a number of countries have shared in this development, American advances in techniques and applications of sampling have been the most striking—the more to be valued because they have been so fully described in numerous papers in the technical journals, more especially the *J.A.S.A.*

In Great Britain, the situation is rather different. Not only is the use of sampling more limited than in the United States but, for various reasons, the surveys that are carried out are rarely published. The result is that statisticians elsewhere have little opportunity of becoming familiar with the fields in which sampling is used in this country, and the methods employed.

The reports issued in conjunction with the meetings of the U.N.O. Sub-Commission on Statistical Sampling [1] indicate both the considerable use that is being made of sampling in many countries and the need for widespread collection of data and exchange of ideas on sampling methods. It is the purpose of this article to describe and comment on the more important sampling investigations carried out in this country by Government departments, commercial organisations and other bodies; and to discuss the main features of English sampling practice, pointing out deficiencies and possible developments in the use of sampling here. Discussion will be confined to investigations concerned with human populations leaving out of account, for instance, research carried out on agricultural experimentation, where sampling methods of a far more involved and refined character are employed. Throughout

* My thanks are due to statisticians in Government departments and elsewhere for making available to me much of the information on which this article is based.

the article, emphasis is on the fields of application and sampling procedures, rather than the results of surveys.

THE DEVELOPMENT OF SAMPLING IN GREAT BRITAIN

The present state of sampling in this country is best seen against the background of a long drawn out development. The first proper use of sampling techniques was in Professor Bowley's survey of Reading in 1912 [2]. He took approximately every 20th working-class household, paying great attention to the calculation of sampling errors and to the possibility of bias being introduced through substitution and refusals. This pioneering use of sampling by Bowley proved a great stimulus to social surveys, which had hitherto been based on non-random selection (Booth, 1889) or complete enumeration (Rowntree, 1901). All the major surveys of the inter-war years, such as the New London Survey, Merseyside, Southampton and others, follow to a greater or lesser extent the sampling methods first used by Bowley.

The first use of sampling in connection with official information was in John Hilton's enquiry into "The Personal Circumstances and Industrial History of 10,000 Claimants to Unemployment Benefit." Extreme care was taken to achieve a representative sample (the 10,000 workers were approximately a one per cent sample of all the claimants) and to avoid the many possible sources of bias. The sampling procedure was fully described by Professor Hilton [3] and, as F. F. Stephan points out in a recent paper [4], it is odd that his methods were not imitated by other Government departments. It was not, in fact, until 1937 that sampling was used in any large-scale Government investigation. In that year, the Ministry of Labour undertook its enquiry into working class expenditure [5] with the object of furnishing information on which a revised Cost-of-Living Index could be based. Budgets of expenditure for four weeks (spaced at quarterly intervals) were obtained from about 10,000 households—the initial selection of households having been made by sampling at random from various registers. The planned revision of the Cost-of-Living Index was interrupted by the war, but the new Interim Index of Retail Prices, which has taken the place of the old Index, is based on the results of the 1937/8 survey [6].

The 'thirties also saw the beginning of Listener Research and other Opinion Research bodies, but it was the war which, as in America, gave the decisive impetus to the utilization of sampling techniques. The need for quick and cheap information led to the foundation of an official Social Survey Unit and to other projects which will be described in this article.

THE NATIONAL LISTS

Perhaps the most important single difference between British and American sampling practice lies in the existence, in this country, of several lists covering the population; certainly, this is the key to what may appear to be simply lack of enterprise—our failure, so far, to make more use of modern sampling developments, and particularly area sampling. It will be seen that nearly all the major surveys to be described are based on samples selected from one or other of the lists. For that reason, and because the coverage and accuracy of the lists differ, a short note on each of them will be useful at this point.

(1) *The Maintenance Register*

This national register started in September 1939, when National Registration came into force. It is kept at local Food Offices and covers the whole population, with the exception of the Armed Forces and Seamen. The Register is a live one, in so far as cards of persons, who have moved out of or into a district are refiled fairly rapidly in the new district. There are separate files for persons under 16 and 16 and over, the cards being filed in order of code number (depending on the district where the card was initially issued).

Apart from the code-numbers, the card gives details of the person's name, sex, address and age (last birthday and date of birth). It must be said that not all the cards have been filled in with all these details and there is a certain amount of inaccuracy; but efforts are being made to bring the Register up to a higher standard of accuracy and completeness.

It will be seen that it is simple to draw a random sample of under 16's or 16's and over, or any other large age and sex group from this register.

(2) *The Ministry of Food Files*

The civilian Ration Books in this country last for a period of one year and there is an exchange of new for old books every July. At this time, the reference pages are extracted from the old books and are filed in alphabetical order of surnames at the local Food Offices. Unlike the Maintenance Register, the Ministry of Food file is only brought up to date once a year so that at any time during the year, the Register in any particular area will include persons who have died or have moved out of that area since last July. Dummy pages are inserted for persons moving into the area. The Register is divided into three separate groups:—under 5, 5–18, over 18.

(3) *The Rating List*[1]

The Rating ledger, kept by the Rating Officers of local authorities, is a list of all the rateable units in the area. It is generally in order of wards

[1] "Rates" correspond to U. S. local taxes.

and in alphabetical street order within wards; within streets, the order is simply by street-number. The entry for each dwelling unit (and, for rating purposes, flats are separate units) gives a description of the property (house, flat, cinema, shop, etc.) and the names of both the owner and occupier. As these lists are used for rating purposes, they are fairly accurate and up-to-date. By eliminating those types of property not required, it is clearly possible to form a sample of dwelling units. In order to sample successfully for households from this list, allowance must be made for the existence of multiple households and the difficulties arising from the distinction between households and dwelling-units.

(4) *The Electoral Rolls*

This is a list, published annually, of all persons entitled to vote in elections; that is, broadly speaking, British subjects aged 21 and over. It is arranged in the same order as the Rating List but, of course, gives not only the occupier of, say, a flat, but all persons in the flat who are eligible to vote. The pre-war list of 1939 was especially useful, in that the head of the household and his wife were indicated by separate symbols. On the new register, this distinction is not made. This register is not very accurate and, if employed at all for sampling purposes, it should be used in conjunction with and checked against the Rating List.

BRITISH SAMPLE SURVEYS

A discussion of sampling surveys falls conveniently into three parts. In the first place, a fairly full account will be given of official sampling investigations; then, some of the more important surveys undertaken by semi-public bodies and research organizations will be mentioned. Finally, an indication will be given of the work of Market Research and Public Opinion bodies.

A

OFFICIAL SAMPLE SURVEYS

(1) *The Social Survey*

The most important organisation in the Sampling Survey field is the Government Social Survey Unit. Founded in 1941, to undertake surveys urgently needed in the administration of Britain's war economy, it is now (as part of the Central Office of Information) a well-established Government research organization receiving an annual Treasury[2] grant of £60,000. The last two or three years have seen not only a considerable expansion in its size and its output—the Social Survey now employs some 250 field workers distributed throughout the country and a research and administrative staff of about 80—but also an im-

[2] The functions of H. M. Treasury are considerably wider than those of the U. S. Treasury Dept.

provement in the quality of its work, including its sampling techniques.

The position of the Survey unit in Government administration is best explained by indicating the procedure governing its work. A survey is planned in response to a request from a Government department, but is put into the field only after Treasury approval has been given. When the survey is finished, a report stating its methods and results is written by the Research Officer in charge, and is submitted to the department concerned for interpretation. While this procedure has the advantage that the Survey undertakes surveys covering a wide variety of subjects, it appears to involve certain drawbacks which are worthy of mention. These are essentially drawbacks arising out of the position, rather than the work, of the Survey.

a. The work of the Survey is largely confined to *ad hoc* investigations, the results of which are urgently required by Government departments. Consequently, not enough time is left for surveys which may be of interest more from a sociological than from a direct administrative point of view. More especially, as it is difficult to experiment with methods in a survey the results of which have been requested and are to be used by a department, there is not as much methodological research as appears desirable. It is to be hoped that the Social Survey will be able to spend more and more of its time on research into sampling techniques, interviewing methods, questionnaire biasses and all the other problems associated with social surveys.

b. The position regarding the publication of the results of the Social Survey is highly unsatisfactory. The reports are, of course, sent to the relevant department and many of them are made available on request at the offices of the Survey. The Survey of Sickness results are published by the Registrar-General and, every now and then, the results of a survey find their way into the Press. It is essential that the reports be given a wider circulation. They might, for instance, be offered for sale at H.M.S.O.

c. Thirdly, it is difficult to escape the impression that the conducting of the survey and the interpreting of its results are too widely separated. The Social Survey ought to be more than a mere collecting agency. The Government department usually keeps well in touch during the planning and execution of a survey but there should be more consultation between the department and the Research Officer in charge of the particular survey over the interpretation of the results and any actions arising directly from them.

It is, of course, impossible in this article to refer to many of the 150 or so sample surveys which the Social Survey has undertaken but it may be useful to name some of the more interesting investigations conducted for different departments and then to say something about the sampling techniques employed. The following list gives an idea of the scope of the Survey's work:—

For the *Board of Trade*

Numerous surveys on wartime shortages of consumer goods; on effects of and attitudes to clothes rationing; on the use of demobilisation coupons; on public knowledge of the need for an active export policy, etc.

For the *Ministry of Food*

Numerous surveys on different aspects of rationing; on the attitude to National wheatmeal bread and to National Milk and Cocoa, etc.

For the *Ministry of Information*

Surveys on the public attitude to various films, books, posters and other publicity media.

For the *Ministry of Labour.*

Surveys connected with the recruitment of women for industry; of miners and of agricultural workers.

For the *Ministry of Health.*

The Survey of Sickness; surveys on the public reaction to the publicity campaign on V.D. and diphtheria immunisation; and on the public attitude to the nursing profession.

For the *Ministry of War Transport.*

Surveys on workers' transport difficulties and on road safety.

In addition there have been regional surveys in Middlesbrough and Willesden; a survey on the impact of air raids; a large survey designed to give a picture of the distribution of the population and of the family and household composition throughout the country, surveys on the demand for holidays, shopping hours, the incidence of deafness, the employment of old persons and many more.

It is evident from this selection that the surveys undertaken by the Social Survey not only cover a very wide field, but also that they concern a large variety of populations, so that sampling methods vary a great deal. According to Box and Thomas [7]

> "The types of population sampled in Wartime Social Survey inquiries may be classified roughly into three groups. In many inquiries, information is required about the whole adult civilian population of Great Britain. A rather larger group of problems concerns only housewives . . . . A third type of problem relates to particular groups."[3]

A fairly typical Social Survey sampling scheme is that used in its

[3] It would now be more accurate to say that most of the surveys concern the adult civilian population.

Survey of Sickness. This regular monthly survey, which began in 1945, is an attempt to derive information about the incidence of all kinds of illness in the adult population. Each monthly sample of about 3000 adults (over 16) is obtained as follows:—The correct proportion of interviews is allocated (according to the current population estimate of the Registrar-General) to each of the twelve Civil Defence Regions. Within each Civil Defence Region, the administrative districts are divided into three groups:—

(a) Towns large enough (over 300,000) to be entitled to 30 or more interviews. (The Social Survey tries to arrange its samples so that about 30 interviews are allocated to each interviewer. This is found to be necessary from the point of view of cost and interviewer efficiency.)
(b) Other Towns.
(c) Rural Districts.

In Group (a), no further division is made. The Registers of all the towns are sampled (such towns account for about 10 per cent of the national sample). In Group (b), the towns are divided into groups by population size in such a manner that each size group is entitled to about 30 (or multiples of 30) interviews. The rural entitlement of interviews for the region is divided into units of approximately 30 interviews each. In both group (b) and (c), the requisite number of towns and of rural districts is chosen, as far as possible, at random, consideration being given to ensure that each of a number of geographical sub-regions receives approximately the correct quota of interviews. In all cases, the individuals for interview are taken at regular intervals from the Maintenance Register in the particular towns and rural areas selected. The over-all sampling fraction is about 1 in 10,000.

In other surveys, such as the investigation of water-heating appliances in domestic dwellings or that of crockery stocks, interest is in households rather than individuals. In these surveys, the samples were selected from the Electoral Rolls and the Rating List respectively. In yet other surveys, such as those on Road Safety, the Demand for Holidays and that on Shopping Hours, quota sampling was used.

It will be seen that so much variation in sampling practice exists within the Social Survey that it is difficult to give a complete picture. The following general points may, however, be made:—

a. In the surveys concerned with the adult population of Great Britain, the samples are usually between 1500 and 3000 cases.

b. For most of the surveys, one or other of the lists mentioned above form the basis of the sampling. The Maintenance Register and the Rating

List are the most frequently used, for the sampling of individuals and households respectively. In surveys connected with special populations, recourse is made to other records. So, for instance, in the Survey of Pneumokoniosis, the sample of 900 was taken from Ministry of Fuel records of men certified by the Silicosis Boards; in the survey on Blood Transfusion, the sample of 1200 donors was taken from local records of the Blood Transfusion Service.

c. The weak link in sampling schemes such as the one used in the Survey of Sickness lies in the selection of the actual towns and rural districts; as indicated above, this selection is largely governed by the desire to achieve complete geographical coverage of the area. But it is limited by considerations of cost and interviewer economy and efficiency. The samples are small, in terms of absolute numbers, and the attempt to allocate the interviews in bunches of at least 30 necessitates principles of selection which are not altogether satisfactory. It is believed that the Social Survey is giving thought to this problem.

d. It is satisfactory to note that the Social Survey is moving further and further away from the use of quota sampling which, at one time, was its chief method of sampling. It may be hoped that, ultimately, the method may be abandoned altogether.

The discussion of the work of the Social Survey has, inevitably, been short and incomplete. There is no doubt that the Survey has done very useful work and also, as its officers themselves would probably agree, that there is plenty of room for improvement in its methods. The time is ripe for a complete and up-to-date statement of the Survey's work on the lines of the paper presented in 1944 by Box and Thomas and, perhaps even more urgently, for a full and critical examination of all the techniques employed in its surveys.

(2) *The Ministry of Food*

The Ministry of Food was one of the first Government Departments to make any considerable use of sampling surveys. In addition to surveys undertaken for the Ministry by commercial organisations, the Social Survey has, during the last few years, carried out a large number of ad hoc investigations on different aspects of the country's food situation. The most important of the Ministry's various survey projects, however, is its continuous Family Food Survey, the aims and methods of which will now be outlined.

Object. The Family Food Survey is the chief source of information concerning the dietary habits of the population. Started in 1941 its objects are (a) to investigate "the nutritional value of food actually consumed as compared with the estimated physiological needs of the same families"; and (b) to collect data on the uptake of welfare foods

and on the effect of food control measures, including price changes. A household or family is defined as "all persons for whom the housewife caters".

Population. From the beginning of the survey the main sample has concentrated on households representative of the working class population of the country. In addition, similar middle class and rural working class samples have been taken more or less continuously for comparative purposes in most years. Special groups, such as miners, agricultural workers and old age pensioners are investigated from time to time.

The Sample. The Ministry take a fresh sample of approximately 800 households every month, so that about 10,000 households are covered annually. The sampling is done in two stages.

(1) Towns, including all the great conurbations, are first chosen by purposive selection—i.e. with regard to their size and the character of the region.

(2) Within the selected towns, sampling is done at random from the Electoral Register. In practice, the procedure is to tackle one part of the town, or ward, at a time so that interviewing is done as economically as possible. Predominantly upper class wards are excluded from the sample, as are those consisting mainly of industrial premises. Any middle-class households sampled in mixed wards are added into a separate middle-class sample. When lists for a whole ward are exhausted (the sampling fraction is 1 in 35), a new ward is started on. This continues until the whole town is covered; then a new and, if possible, similar town is chosen for sampling.

The Ministry of Food points out that the representativeness of the resultant sample is subject to two qualifications:

a. In the first place, small towns tend to be slightly under-represented, mainly as a result of the inevitable immobility of the interviewers.

b. In the second place, in the past about 20 per cent of the households sampled proved to be non-contacts even after three calls, while a further 30 per cent could not or would not co-operate for different reasons. When all else fails, interviewers are permitted to take substitutes by calling at each house in turn to the right of the one originally sampled.

This large-scale substitution is a potential source of bias and this must be borne in mind in interpreting the results. The fact that the sample has had a somewhat overweighted proportion of children may be attributable to this substitution. As far as average food expenditure per head is concerned, analysis has shown little difference between pre-selected and substituted households.

The fieldwork of the survey is undertaken by an outside commercial agency, so that direct contact between the Ministry and informant is avoided. Fieldworkers call every second day during the week of investigation or, if necessary, every day.

Sampling Variation. Another point to consider is the variability of the population in terms of family size, sex and age composition and the number of meals taken at home and outside. For this reason, as the Ministry points out, repeat samples taken at the same time and with the same regional distribution will vary slightly. Percentage coefficients of variation have been calculated, for all the different food expenditures, nutrient intakes, etc.

Procedure and Data Obtained. Each housewife who agrees to cooperate, is given a log book to complete for the succeeding seven days. For each day, she has to fill in four sections as follows:—

(1) Quantity, description and cost of each item of food bought on that day.
(2) Quantity, description, source of and price paid (if any) for all home-grown food, gifts and welfare food.
(3) Description of food served at each meal and note indicating which persons were present.
(4) Number and type of meals taken out by different members of the household.

In addition, the interviewer records on the cover page, details of family composition, including age, sex, relationship to housewife and occupation of all members of the household. This information makes the calculation of the nutritional needs of the family possible. Furthermore, food in the larder is weighed and measured at the beginning and end of the week, so that changes in stocks may be calculated.

Results. The results, calculated for the sample every month, show:

a. The average daily consumption per head of each food.
b. The average daily intake per head, and average daily requirement per head of each nutrient.
c. The average expenditure on and actual price paid for each food, and on all foods. Also the quantity and value of foods obtained from free sources.

(3) *The Ministry of Labour*

The Ministry of Labour has made less use of sampling than might perhaps have been expected. Reference has already been made to its 1937/8 Family Budget enquiry and to Professor Hilton's investigation. The other use of sampling which should be mentioned here is in the

Ministry's Analysis by Age-group of Insured Workers. Up to July 1948, at the time of the annual exchange of unemployment books, information about the distribution of insured workers by industry, sex and by four age-groups (under 16, 16 and 17, 18-20, 21 and over) was obtained. It was desired to obtain more detailed information for the 21 and over age-group and this was done first in 1937 and regularly since 1942 by means of sampling.

The sample was taken from bundles of about 100 unemployment books tied up at the Employment Exchanges (for dispatch to another office). Two books were selected out of every bundle, respectively $\frac{1}{3}$ and $\frac{2}{3}$ of the way through the bundle. The resultant 2 per cent sample (about 300,000) was a small one and the Ministry of Labour pointed out that great care needed to be taken in interpreting results.

These samples yielded analyses for total insured workers by sex, age and industry and by five-year age-groups. Up to 1947, the results are available only on a national basis; in that year, for the first time, a regional analysis was also published.

(4) *The Ministry of National Insurance*

The analysis based on the annual exchange of all unemployment books (not the additional sample analysis) not only yielded detailed information about the insured working population, but also served as a basis for estimates of the total working population.

With the new National Inusrance Act, which came into force on 5 July, 1948, a different type of insurance book is used and their large number (25,000,000) renders a simultaneous exchange of books impracticable.

The exchange of books is consequently being spread over the year at quarterly intervals and a careful sampling plan has had to be devised in order that detailed estimates for the whole working population may be made.

The books have been allocated to the four quarters in a systematic and unbiassed manner. No details have so far been announced about the kind of estimates it is intended to make at the quarterly dates, or the way in which they will be combined to give annual estimates.

(5) *The Ministry of Works*

The Ministry of Works uses two methods for obtaining details of the Labour Force in the Building Industry. In the first place, a quarterly census of the whole industry is taken yielding very detailed data on the distribution of the labour force by region, occupation, type of

work, etc. Owing to the large number of forms involved and the delay in getting them in, the first results are not available until 12 weeks after the Census date.

The other source of information is a monthly sample (started in 1945), for which the population consists only of the twelve main building trades (about $\frac{2}{3}$ of the whole industry). Information collected from the sample is confined to the total number of operatives employed by firms. Against these disadvantages of more limited coverage and less detailed information of the sample enquiry must be set the great economy for both the Ministry and the industry, increase in accuracy and the fact that the final tabulation is available three weeks after the sample date.

The Ministry of Works requires that the maximum error in the over-all total of operatives should be 1 per cent and found that the appropriate sample size would be between 5000 and 6000 contractors. The sample is based on the last available census tabulations (usually from a census taken 5–7 months previously). The sampling fraction varies from 1 in 100 for firms with 0 employees, 1 in 30 for firms with 1–5 employees to 1 in 2 for firms with 71–99 employees and a 100 per cent sample for all firms with 100 and more employees. The over-all sampling fraction is approximately 1 in 20, there being about 120,000 firms in the 12 Trades with which this sample is concerned.

The sampling in these investigations of the Ministry of Works is designed and executed with particular care and the possibility of taking a census less frequently and deriving more detailed information from the samples is being investigated.

(6) *The General Register Office*

The General Register Office is the department responsible for the Decennial Censuses of Population and for the collection and publication of demographic statistics generally. In view of the scope of its work, its failure so far to make use of sampling to any considerable extent is striking. There are two uses of sampling by the G.R.O. to which reference may be made:

(1) "Classification and tabulation of multiple or secondary causes of death" [8]. In the classification of death by cause, when more than one cause is mentioned, it is necessary to select one as that to which the death should be classed. Information should, however, be collected not only on the frequency of occurrence of each cause as a primary or secondary cause but also on the frequency with which certain causes appear in conjunction.

In each of the years 1921–1930, a group of causes of death was selected for further investigation. Apart from the ordinary punching of the primary cause of death on each card, the whole cause as certified was written at the top of the card. A sample was then taken for each group of deaths attributed to a particular primary cause; the sampling fraction varying from 1 in 10 for groups with a large number of deaths to a 100 per cent sample for groups with only small numbers.

The secondary causes of death on all the sampled cards were then coded and an analysis of associated causes was made.

(2) "Emergency Medical Service Records." Records are available for all patients who have been treated as in-patients under the E.M.S. A 1 in 5 sample of all the cards of discharged patients was taken, yielding about 45,000 cases for each year from the beginning of the war to the end of 1947. The data collected and tabulated from this survey included sex, age, civil status, branch of service or other occupation, and full details of the patient's hospital record, from his admission to his discharge.

It will be seen that both the above samples were samples from documents and the only field sample survey with which the General Register Office has, in fact, been in any way associated is the Survey of Sickness (mentioned in Section 1) the results of which are now published in the General Register Office Quarterly Returns.

Some general explanations of the apparent lack of enterprise regarding the use of sampling in Government departments will be offered in a later Section. It is, however, worth noting at this point that the Registrar-General is considering the use of sampling at the 1951 Census of Population but that no conclusions have so far been reached.

There is an unanswerable case for the use of sampling at the next Census to get out preliminary results more quickly. Furthermore there is no reason why all the final Census tabulations should be based on an anlysis of *all* the returns. Supposing that the same information was obtained from every member of the population (thus avoiding any legal difficulties), results of sufficient accuracy might be obtained on a number of questions by taking samples of the returns. This would save considerable time, labour and money and would avoid the situation arising out of the 1931 Census, some of the results of which it has still not been possible to analyse.

It is to be hoped that the Registrar-General will decide in favour of sampling and that no time will be lost in undertaking the necessary research and other preparations.

(7) *The Sample Family Census*

The Sample Family Census, which took place at the beginning of 1946, is an essential part of the work of the Royal Commission on Population. Its purpose was to obtain information about changes in the size of families and thus to provide data of obvious importance with regard to "the population problem and its bearing on housing, family allowances, social insurance, and other measures of social welfare" [9].

The sample of 1 in 10 of all the married women in the country (yielding about 1,600,000 women) was selected from the Food Office files. At the 1945 exchange of ration books, women had been asked to describe themselves as "Miss" or "Mrs." Accordingly, after every 10th reference page had been extracted from the file of persons aged 18 and over, the pages of all males and of all females describing themselves as "Miss" were discarded (apart from women who, though they described themselves as "Miss," had changed their names since the last exchange of ration books). The remainder then, were women who had described themselves as "Mrs." or who had not given the information. In order to preserve the randomness of the sample, *all* these women were contacted and those who were actually found to be "Miss" were then discarded from the sample. One more precaution was necessary before this could be regarded as the final sample of presently or formerly married women: the pages for women who had left the particular area were removed by checking the reference pages against the Maintenance Register. Great efforts were made to contact all the women sampled and substitutes were not permitted in any circumstances.

Questions were asked on:—whether at present married, widowed or divorced; date of birth and of first marriage (and of the end of the marriage, if applicable); dates of birth of all live-born children; number of children under 16 alive, and husband's occupation.

The information was collected by 12,000 enumerators who received 1/4d for each completed form. It is hoped to publish the report on the Census, with full details of the sampling methods employed and accuracy achieved, sometime next year.

(8) *The National Farm Survey*

The National Farm Survey was a development from the local farm surveys which were being carried out during the war to assist the work of the County War Agricultural Committees. The task of these Committees "may be shortly stated as ensuring that each farm makes its maximum contribution to food production" [10].

The National Farm Survey was carried out from 1941 to 1943 and consisted, in the main, of information obtained according to a *standardised* pattern from the local surveys and of the returns from the 1941 agricultural Census. The survey, in addition to aiding the local committees, provided invaluable statistical material on a national and county basis.

The Survey population consisted of the 300,000 holdings of 5 acres and over. Information was collected on the following subjects:—

A. Type of tenure, rent and length of occupation.
B. Economic type of occupier.
C. Size and type of holding.
D. Convenience of farm lay-out; situation of holding; condition of buildings.
E. Nature of the soil; nature of water and electricity supply.
F. Managerial efficiency of occupier and condition of cultivated land.

It was impracticable and unnecessary to base the national and county totals and averages on an analysis of the records of all the 300,000 holdings. A random sample was therefore drawn and was stratified by county and size of holding. The sampling fractions varied according to size of holding as follows:

Size of Holding Acres	Sampling Fraction %
5– 24.9	5
25– 99.9	10
100–299.9	25
300–699.9	50
700 and over	100

The final sample covered about 14 per cent of all the holdings in England and Wales. The figures throughout the report were arrived at by multiplying the sample results by the reciprocal of the respective sampling fractions. The full published report includes a detailed description of the sample design, the standard errors (which are trivial for national and very small for county data) and estimates of the gain through (a) stratification and (b) variable sampling fraction.

(9) *BBC Listener Research*

It is appropriate, at this stage, to say something about the work of the Listener Research Department of The British Broadcasting Corporation. This department, first set up in 1936, collects information

about the listening habits and tastes of the British public; it tries, in fact, to establish some kind of a "box-office" substitute in the world of radio. The head of the department has described the two principal tasks of Listener Research as being:—

(1) to find out how many people listen to each broadcast,
(2) to ascertain listeners' opinions of the broadcasts which they hear [11].

To accomplish these tasks, Listener Research organizes two separate and very different enquiries, which will now be described in a little detail.

The Survey of Listening

This continuous survey is based on a daily sample of 3000 persons (aged 16 and over). An equal number is interviewed in each of the six BBC broadcasting regions into which the country is divided, but the results are subsequently weighted according to the population in each region. A further sub-division into rural and urban areas is made within regions. Selection of individuals in each area is by means of quota sampling—according to quotas based on the sex, age and social class distribution of the population. The procedure employed for determining the social class of the informant is for the interviewer to ascertain his occupation and with the aid of a "common-sense assessment of the contact's manner, bearing and conversation" to classify him as working, lower middle or upper middle class. It may be questioned whether this procedure is reliable and, considering the relevance of social class in this type of survey, one might suggest the use of a more precise method.

In the interview, the contact is asked to name the programmes he listened to on the previous day. The interviewer will try to aid the informant in remembering the broadcasts and may have to distinguish between "listening to" and simply "hearing" a programme. This, of course, is the great snag of all radio research and it is doubtful whether there is any satisfactory solution.

The results of these daily interviews are issued about 7 days after the broadcast in the form of what is called the Listening Barometer; that is to say, against each programme are shown the percentages of the sample (in each region and in Great Britain as a whole) who listened to it.

The Listening Panel

While it is of great value to the BBC to know the number of people who listened to each broadcast, it is equally important to know how

a particular broadcast was liked by those who did actually listen to it. For this purpose, the BBC has established a panel made up of volunteers recruited as a result of occasional appeals. This panel consists of 3600 members, distributed equally over the six regions. Each panel member receives, twice a week, a questionnaire containing questions about a number of broadcasts to be made during the next few days. All types of programmes are included and panel members are specifically asked to give their opinions only on those broadcasts in which they were anyhow interested; that is to say, duty listening is to be avoided. The panel members (whose average "panel life" is 18–24 months) are, of course, self-selected and it is questionable whether the opinions sent in on a particular broadcast can be regarded as fully representative of the opinions of all who listened to it. The summarised results of the questionnaires are passed on to the producers of the programmes.

The results of Listener Research are given the widest possible circulation inside the BBC, but it is impossible to assess the extent to which they influence programme policy and planning. As Listener Research constitutes the main link between the BBC and its public, directors of sections and programme producers probably take a considerable interest in its results.

The above account of the use of sampling in official surveys cannot be regarded as complete. It has not been possible, for instance, to say anything about the use of sampling in the Ministry of Home Security's wartime investigations of the effects of bombing; or of the sample enquiry which the Royal Commission on the Press is believed to be making; or for that matter, of the employment of sampling methods in the internal work of Government departments. It is believed, however, that enough has been said about the most important official sample surveys, for which any details could be obtained, to form a basis for the general remarks which follow in the last sections of this paper.

B

SURVEYS CONDUCTED BY SEMI-PUBLIC BODIES AND RESEARCH INSTITUTES

An indication has already been given of the wide use of sampling in social surveys and there is an increasing tendency, on the part of Universities, Research Institutes and local authorities, to avail themselves of sampling techniques in their surveys of social and economic problems. It is now proposed to describe a few of the more important and interesting surveys which have recently been, or are still being, carried out in this field by non-official, non-commercial organizations.

1. *Population Investigation Committee—Survey of Maternity*

The survey on maternity [12] was undertaken in order to obtain data on the social and economic aspects of childbearing. In particular, information was sought on the availability and use made of maternity services in different parts of the country and social classes, and on present-day expenditure on childbirth.

It was desired to collect the information through personal interviews with mothers "whose experience could be regarded as typical of the whole population of women now bearing children." The method adopted was to take a sample in time *from all* local authorities—rather than a sample *of* local authorities—and to interview *all* women who were delivered in England, Wales and Scotland during the week of 3–9 March, 1946. This procedure had the advantage of greatly easing and speeding up both the field work and the administration of the survey.

Officials of all the 458 local authorities were approached and active support was received from 424 (92%) of them. It is estimated that 16,695 births were registered during the week in question, of which 15,130 were notified to the Survey Committee. Successful interviews were made regarding 13,687 (or 90.5%) of these. It is interesting to note that refusals were obtained in only 2% of the cases.

As a great deal of information had to be asked for, it was decided to issue two separate types of questionnaires. Type 'A' dealt mainly with the use made of maternity services, while Type 'B' concentrated on the financial aspects. Certain basic questions were, of course, asked in both. The 424 authorities were split at random into two groups, one group for the Type 'A' and the other for the Type 'B' questionnaire. The proportion of successful interviews was 91.5% for Type 'A' and 89.3% for Type 'B' questionnaire. Analysis shows that, in most important respects, the samples of mothers making up the two surveys are closely comparable.

The results of the questionnaire inquiry—which, together with the methods, are very fully described in the report—are supplemented by detailed studies of the extent and quality of maternity services in selected areas.

One aspect of particular interest remains to be noted. It was decided to conduct a follow-up survey in order to study

(a) differences in morbidity of different groups of mothers and children;
(b) the factors which influence the health and development of different groups of children during their first two years of life;
(c) differences in morbidity and development of full-time and premature children.

It was felt to be unnecessary to apply the questionnaire to all the mothers in the original enquiry, as long as the absolute number in each social group was large enough to yield the precision required. For (a) and (b), the sample consisted of one out of four mothers and children in the "manual workers" group (chosen at random) and all those in the other groups. This gave a total of about 5000 cases.

This follow-up survey took place in March 1948 and met with considerable success. Special efforts were made to trace those women who had left their original address and completed questionnaires have been received from about 94% of the women.

For the survey concerning premature children, a different method of selection was employed. Each of the 800 or so premature babies in the original sample was matched with a full-time one of the same sex, social class and birth order and born to a mother of the same age, living in the same region and at the same room density. A number of the babies subsequently died while others moved out of the original area, so that a good number of the pairs had to be abandoned. It has, however, proved possible to follow-up 640 pairs through their first two years of life. This unusually large-scale and carefully conducted controlled investigation will, in fact, provide invaluable material on the causes of prematurity; on possible ways of preventing deaths arising from it; and on the question as to the extent to which premature and full-time babies differ in their subsequent development.

2. *Population Investigation Committee—Inquiries into the Trend of Intelligence*

In 1947, the Population Investigation Committee and the Scottish Council for Research in Education jointly initiated an inquiry into changes in national intelligence in Scotland. The purpose of the inquiry is to test the hypothesis that the "current patterns of differential fertility (which show a negative correlation between size of family and the measured intelligence of the children) really imply a fall in national intelligence."

The present inquiry has been conducted in Scotland, because in 1932 a complete age-group of Scottish children had been the subject of an inquiry into intelligence and it seemed appropriate to apply the same tests, which had then been used, under similar conditions to the same *age-group* of children. The survey has covered some 80,000 children born in 1936, in respect of whom the same group intelligence test has been administered and for whom an individual questionnaire was completed, giving details of age, sex, size of and position in family, school and class, etc.

A special and more detailed questionnaire has been filled in by a random sample of children, consisting of all children born on the first three days of each month in 1936 ("36-day sample"). Furthermore, individual Binet tests have been administered to a further sub-sample consisting of all children born on the first day of each alternate month ("6-day sample").

The field work has been highly successful. The group tests were administered to 88% of the children and a greater number completed the questionnaire; the more detailed questionnaire was obtained from 99% of the 36-day sample, while the Binet tests were given to 99% of the 6-day sample.

Results, which are not yet available, will show the change in the average and standard deviation of the I.Q. since 1932; and the relationship between I Q. and size of family, parental occupation and other factors. It is intended to follow up a sample of the children, in order to compare the development of children of different I.Q.'s over a number of years. For this purpose, the 6-day sample will be taken plus 400 children of very high I Q. and 400 children of very low I.Q.

3. *London School of Economics Social Research Division—Survey on Social Class and Social Mobility*

The aim of this survey, financed by the Nuffield Trust, is to discover what people mean by social class, and what are the chief factors accounting for class differences and for the movement from one class to another. An attempt will be made to estimate the relative importance of these factors and to trace the changes which have taken place in their influence over past years.

A start has been made with a factual survey carried out in two London Boroughs and a neighbouring rural area, designed to test a method of obtaining information about educational opportunity and occupations entered in successive generations. The possibility of extending this investigation to other parts of the country is under consideration.

The aim in this survey was to interview a married person, preferably the wife, and the introduction to the married person was obtained by drawing a random sample of individuals from all those of age 16 and over resident in the area investigated. If the individual drawn was unmarried, but living with parents, the mother was to be interviewed. If the individual drawn was of age 60 or over living with married son or daughter, the young wife was to be interviewed. Persons not to be interviewed were:

(a) Unmarried persons living apart from their parents.
(b) Persons previously married (to avoid complications possibly arising from differential treatment of step-children).
(c) Separated or widowed persons.

It is expected that this research will extend over approximately five years.

4. *Local Social Surveys*

The early social surveys, to some of which reference has already been made, were primarily focused on the central problem of poverty and its different aspects. In the late '30's, the emphasis began to shift. Attention was being directed more and more to questions connected with town planning and life on housing estates as, for instance, in the surveys of Birmingham [13], Becontree [14] and Watling [15]. Since then, and especially since the end of the war, a large number of regional planning surveys have been initiated and published. The social survey has rightly become an integral part of town planning and most of the recent planning studies will be found to contain a social and economic survey, usually based on a sample of dwellings or households. Thus, for instance, the Middlesbrough Plan [16] includes a sample survey as do the plans of Luton [17], Worcester [18] and some of the studies made by the Association for Planning and Regional Reconstruction. The sampling methods employed are generally not of great intrinsic interest and it is proposed here to describe, as an example, only the survey of Luton, which is one of the best published so far.

Report on Luton (1945)

Luton, described as a "young, expanding industrial town" surrounded by rural areas, lies 20 miles from the northern fringe of London and has approximately 100,000 inhabitants. It was felt that, in view of current housing problems and the present-day changes in education, health and other services, the Local Council should have up-to-date information on population, housing and the social aspects of public services.

As the authors were primarily interested in population and housing, they would have liked to survey every occupied dwelling. This, however was impossible and a sample of dwellings was taken. Luton was divided into three areas. District I—areas of unfit houses; District II—the older parts of Luton; District III—the rest of the Borough. The population density of these districts is shown by the average number of

persons per acre, which were 92, 36 and 11 respectively. The sampling fraction varied as follows:

District	No of Houses	Sampling Fraction	No. of Houses Surveyed
I	890	1/1	890
II	5,270	1/5	1,054
III	22,300	1/10	2,230
Total	28,460		4,174

To obtain data for the whole Borough, a sample of 1 in 10 of all dwellings in the borough was achieved by using 1 in 10 of the houses sampled in District I, every other one in District II and every one in District III. The actual addresses in each district were selected from the Rating List.

The survey provided data of three kinds:—a certain amount of information for each sampled household (e.g. no. of rooms, no. of persons, overcrowding); information for each individual (e.g. age, sex, civil status, occupation, education); and detailed fertility data relating to each married woman. On the basis of this full and carefully obtained information, the authors were able to give invaluable guidance for future planning.

PUBLIC OPINION AND MARKET RESEARCH

It is not intended in this article to give an exhaustive account of the surveys conducted or methods employed in opinion polls and market research in this country. Some of the main organisations in the field will be mentioned and an indication will be given of any interesting characteristics they may possess. The justification for this somewhat sketchy treatment is twofold. In the first place, it is in the nature of the work of, at any rate, the market research bodies that the aims, methods and results of their surveys are made known to their clients rather than to the general public. Anything like an accurate account of their activities would be very hard to give. In the second place, it is very much doubted whether either the methods employed by public opinion polls and market research bodies here or the fields of application to which they direct their attention differ sufficiently from American practice in this field to warrant a full discussion in this article.

(i) *British Institute of Public Opinion*

The B.I.P.O., founded in 1936, is one of the international chain of

Gallup institutes. Opinion polls in this country do not, of course, enjoy anything like the publicity given to polls in America. Thus, the findings of the B.I.P.O. are published by only one newspaper, the daily News Chronicle. Nor could it be said that opinion polls here yet arouse such universal interest or such bitter controversy as in the United States.

The B.I.P.O. has, in the past 12 years, conducted polls on a wide variety of topics and has also published fairly accurate forecasts on the 1945 general election and on numerous bye-elections. Apart from the work concerned purely with opinion polls, B.I.P.O. has, on a number of occasions, co-operated in, or conducted surveys for, other organisations. Thus, it undertook part of the war-time body-weight survey for the Ministry of Food and several surveys for the Board of Trade. The method of sampling used by B.I.P.O. is quota sampling.

(ii) *Mass Observation*

The surveys conducted by Mass Observation are not based on scientific sampling methods. Mass Observation is mentioned here because it has received very wide publicity and aroused considerable interest. Founded in 1937 it has, in a very active decade, published some 14 books and several hundred bulletins and articles.

Its reports are based on two different sources: field surveys and a national panel of respondents, neither of which appear to be based on a randomly selected sample.

(iii) *Market Research Bodies*

As in America, there is in England a large and increasing number of market research organisations. As far as can be judged, from the few leaflets which it has been possible to obtain, all of these organisations claim that scientific sampling methods form the basis of their results and standard error formulae are usually quoted. Not enough is published to permit any evaluation of the methods employed by market research bodies; it can be stated, however, that the method of selection generally employed is quota sampling which is cheap (judged purely in money terms), offers great ease of administration and fieldwork (it by-passes the problem of call-backs), and is, for commercial purposes, considered as sufficiently accurate.

The most important organisations in the field are probably the British Market Research Bureau and Research Services Ltd. The former, among other activities, has done a number of surveys for the Board of Trade and the Ministry of Food. Research Services Ltd., is the new name of what was formerly the Research Department of the London

Press Exchange. It now constitutes a separate company and covers a wider range of activities than other market research bodies. Another company, Attwood (Statistics) Ltd., runs a Consumer Panel, selected on the basis of a random sample.

D

A COMPARISON BETWEEN BRITISH AND AMERICAN SAMPLING PRACTICE

The preceding sections show that investigations on a wide variety of subjects are undertaken by means of sampling, both inside Government departments and by other organisations. Nevertheless, it can hardly be claimed that the available sampling techniques are exploited to the full or that sufficient effort is made to experiment with different and more refined methods. The remainder of this article is devoted to a discussion—based on the foregoing survey—of the general features distinguishing the use of sampling in Great Britain.

It is important, in this discussion, to guard against the error of treating the American and British situation as strictly comparable. The enormous area and widely dispersed population of America gives rise, in the first place, to a much greater *necessity* for sampling and has led to the development of sampling methods particularly suited to these circumstances. In Great Britain, the population is relatively small and highly concentrated. It would be a mistake to believe that American sampling methods could necessarily be used with the same success here, and much experiment and research is needed to decide which refinements and developments can usefully be "imported" (it is interesting to compare a French point of view on this subject [19]).

Further to this, it should be pointed out that probably, in this country relatively more data about the population comes in as a by-product of administration and does not need to be obtained by special surveys. (Information about the Labour Force is a case in point.) This difference again implies that the *necessity* for sampling is not usually as urgent as in the United States, but it also points to far greater *opportunities* in the sampling of documents and returns to obtain quick and regular data to aid administration. It cannot be said that anything like full use is being made of these opportunities.

(a) *The Attitude to Sampling in Great Britain*

It must, in the first place, be emphasised that in England sampling is not yet as *generally* understood or accepted as a reliable way of collecting information as in the United States. The American public

was made "sample conscious" by the evident success and through the enormous publicity of public opinion polls in general and their past election forecasts in particular. The corresponding absence of such publicity here partly explains any lack of confidence in sampling felt by the general public.

More important from the point of view of the development of sound sampling practice is the distrust and lack of knowledge remaining, in Government departments and elsewhere, among persons who are potential initiators of sample surveys. No doubt, now that more attention is paid to sampling in University statistics courses, this distrust and ignorance will gradually be dispelled. Yet, there is an urgent need for the spread of knowledge through the full publication of the methods and results of surveys and for a regular procedure by which statisticians and administrators in Government departments and elsewhere may avail themselves of advice on the conduct of sample surveys.

A department, perhaps within the Central Statistical Office, should devote its whole time to the examination of past and current sample surveys in all fields. By scrutinising the methods used in the past and by giving advice on new projects, such a department could help first to establish and ultimately to maintain a high standard of sampling practice. Further, it could act as a central Register of work being done by all sorts of organisations and thus prevent much duplication of research.

(b) *The Publication of Sample Surveys*

It is obvious that the full publication of survey methods and results is a considerable asset to further development. In England, unfortunately, the position regarding publication is very unsatisfactory.

Some indication of the circulation of the Social Survey reports was given earlier; the Family Food Survey of the Ministry of Food, which has continued for seven years and is the main source of information concerning the diet of the British public, is only now receiving publication; as far as can be ascertained, the sample surveys undertaken by the Ministry of Works, the Board of Trade, the Ministry of Labour and the Ministry of Home Security have not been described in any publication.

Non-publication of sampling methods is only one aspect of a general, but possibly decreasing, tendency in Government departments to treat their information as confidential or, at least, not for publication. This is a great barrier to the development of sampling methods and the spread of knowledge. It is to be hoped that the valuable suggestions concerning this problem, made in a recent report by the National

Institute of Economic and Social Research [20], will be followed up. In all fairness, it must be added that there are exceptions to this tendency. Thus, the National Farm Survey was fully described in the official and publicly available report. Similarly, the methods as well as the results of the Sample Family Census will be given in great detail in the forthcoming report on that survey.

There is one more point on this question of publication. With the increasing use of sampling in investigations concerned with human populations, there appears to be a good case for an international journal devoted to this subject. Such a journal would not need to confine itself to sampling, but could range over all the methodological problems which arise in the planning and execution of surveys. Perhaps the United Nations Sub-Commission on Statistical Sampling might consider the question.

(c) *Methodological Research*

The work on sample surveys in this country is largely confined to *ad hoc* surveys or, at any rate, surveys which are important for their results rather than their methods. There is clearly a need for more systematic and coordinated experimentation and research on sampling techniques, on the lines of the work done, for instance, by the Bureau of the Census. Perhaps this task could most suitably be performed by a department such as that suggested in (a) above.

(d) *Area Sampling and Quasi-Random Sampling*

This is not the place to enter into the controversy regarding the respective merits of these two methods of sampling. It is desired merely to indicate why area sampling, which has been so successfully applied in the United States, is not used here; and whether it is likely to be used in the future.

The most obvious reason for the non-use of area sampling here lies in the availability of the National Lists, mentioned earlier. None of these lists is perfectly accurate, but if care is taken, they can form an adequate basis for random sampling at given intervals, i.e., quasi-random sampling. The fact that these lists are available, while maps suitable for area sampling are not (the maps of the Fire Service may be closest to what is required) has acted as a check to experimentation with area sampling methods.

It is probable that even with the relatively small and concentrated population of Great Britain, area sampling would lead to certain

advantages. Research is needed to show whether the high initial outlay would be justified by increased accuracy and economy resulting from its use.

(e) *The Use of Purposive Selection*

It has been seen that, in Great Britain, most of the sample investigations concerned with human populations rely on stratified random sampling from the lists as their method of selection. If national estimates are involved, the sample may have to be spread geographically over the whole country. Often, cost considerations and interviewer resources make it desirable to get the field work reasonably concentrated and, in a number of surveys, the primary sampling units are chosen by purposive selection. Thus, in the Ministry of Food Family Diet Survey, the towns are chosen to be representative of Great Britain, with regard to their size and the character of the region. The same sort of selection is used in some of the Social Survey investigations.

Selection of the primary sampling units on the basis that they are "representative" may easily introduce bias and it will be better to use the alternative procedure of sampling at random from a list of areas or towns. The areas would be grouped in strata in some appropriate way and a variable sampling fraction could be applied, perhaps on the lines used in the United States Sample of the Labour Force [21].

(f) *The Use of Quota Sampling*

Within the primary sampling units, the usual procedure employed for the selection of the final sampling units is quasi-random sampling. It should, however, be noted that quota sampling is still the method generally used by commercial organisations, opinion research bodies (including Listener Research) and occasionally in Social Survey investigations.

The cheapness and ease of execution of quota sampling are qualities which endear it to commercial organisations. But, from a statistical point of view, the method is unsatisfactory for three reasons. In the first place, the choice of units depends on human judgement and, therefore, may involve serious bias. In the second place, quota sampling does not permit any effective control over the work of the interviewers; and finally, it is not possible to attach an estimate of the sampling error to the results. The method should, therefore, not be employed in any surveys on which administrative decisions are to be based.

(g) *Further Uses of Sampling*

There are a number of fields to which sampling might well begin to be applied in this country and there are some indications already that new ground is being broken and that further uses are being contemplated.

The Board of Trade has, at its disposal, a fairly complete list of industrial firms in the country. This might serve as the basis for samples designed to obtain information additional to that supplied by the Censuses of Production and Distribution. It is believed that the lists of firms in specified manufacturing industries are to be utilized for sample surveys on Capital Expenditure.

Much information on family expenditure is needed and will regularly have to be obtained for any contemplated permanent Retail Price Index to take the place of the present interim Index. Sample surveys have been started by the Social Survey with a view to obtaining information on the national expenditure on selected items. So far, surveys on laundry and household textiles have been conducted and ultimately all types of consumer expenditure will be covered.

The most obvious gap in sampling practice in this country is in the field of population. It is to be hoped that the Registrar-General will decide to use sampling in the 1951 Census and that he will also begin to consider the possibility of taking regular sample surveys of the population to fill in the 10-year gaps between Censuses. Sampling for population estimates requires more refined techniques than are customary in most of the surveys mentioned above. It is in this field that American developments have been most striking and any contemplated use of sampling at the next Census of England and Wales should be based on a very careful examination of American experiences.

The use of sampling can lead to great economy in cost, time and personnel and is bound to play an increasingly important role in Government administration and elsewhere. The most urgent needs in this country are for systematic research, in government departments, universities and other bodies, into sampling as well as other (and often more difficult) problems connected with surveys; and for full publication of the methods used in and lessons learnt from completed sample surveys.

REFERENCES

[1] Report of the Sub-Commission on Statistical Sampling (United Nations E/Cn. 3/37) issued 21st October, 1947.
W. E. Deming "A Brief Statement on the Uses of Sampling in Censuses of Population, Agriculture, Public Health and Commerce" published by the United Nations, Lake Success, 1948.

Series of reports submitted to the 2nd meeting of the Sub-Commission at Geneva in September 1948 (United Nations E/Cn. 3/sub. 1/various).

[2] Bowley, A. L., "Working-class Households in Reading," *Journal of the Royal Statistical Society*, 1913.

[3] Hilton, John, "Enquiry by Sample: an Experiment and its Results," *Journal of the Royal Statistical Society*, 1924.

[4] Stephan, F. F., "History of the Uses of Modern Sampling Procedures," Presented at the 25th Session of the International Statistical Institute and reprinted in the *Journal of the American Statistical Association*, Vol. 43, March, 1948.

[5] Ministry of Labour, "Weekly Expenditure of Working-class Households in the U.K., 1937–8," *Ministry of Labour Gazette*, Dec. 1940, Jan. and Feb. 1941.

[6] *Interim Report of the Cost of Living Advisory Committee.* H.M.S.O. Cmd. 7077; and "Index of Retail Prices" *Industrial Relations Handbook 1944*, Supplement No. 2, January, 1948.

[7] Box, K. and Thomas G., "The Wartime Social Survey," *Journal of the Royal Statistical Society*, 1944.

[8] *Registrar-General's Decennial Supplement for England and Wales*, 1931, Part IV (published 1947).

[9] Article on the Sample Family Census in the "*Times.*" January 17th, 1946.

[10] Ministry of Agriculture, "*National Farm Survey of England and Wales,*" A Summary Report. H.M.S.O. 1946.

[11] Silvey, R. J. E., "Methods of Listener Research employed by the B.B.C.," *Journal of the Royal Statistical Society*, 1944.

[12] Population Investigation Committee, "*Maternity in Great Britain,*" Oxford University Press, 1948.

[13] Bourneville Village Trust, "*When We Build Again,*" 1941.

[14] Young, T., "*Becontree and Dagenham,*" 1934.

[15] Durant, Ruth, "*Watling,*" 1939.

[16] Lock, Max and others, "*The County Borough of Middlesbrough Survey and Plan,*" 1947.

[17] Grundy, F. and Titmuss, R. M., "*Report on Luton,*" 1945.

[18] Glaisyer, Janet and others, "*County Town,*" 1946.

[19] Thionet, P., "*Méthodes Statistiques Modernes des Administrations Fédérales aux Etats-Unis,*" Paris, 1946.

[20] National Institute of Economic and Social Research, "*Social and Economic Research and Government Departments,*" 1947.

[21] Hansen, M. H. and Hurwitz, W. N., "*A New Sample of the Population,*" Bureau of the Census, 1944.

UNEMPLOYMENT AND MIGRATION IN THE DEPRESSION (1930–1935)*

RONALD FREEDMAN
AND
AMOS H. HAWLEY
University of Michigan

This is a study of the reciprocal relationship between migration and unemployment in Michigan during the depression period 1930 to 1935. Specifically, it is concerned with two questions:

(1) Do the migrants during a depression have a poor employment history as compared with non-migrants of similar characteristics?

(2) Do the migrants have a better employment experience after migration than non-migrants of similar characteristics?

By the use of matched control groups it is found that the differential in unemployment rates occurs after migration, not before. The results are consistent with the hypothesis that in a depression migrants tend to be at a disadvantage in the new labor market to which they move.

THE PROBLEM

THE QUESTIONS above are at the heart of the more general problem of the relationship between migration and economic opportunity. Comparisons of the pre-migration employment status of the migrants with similar persons who stay at home should indicate whether unemployment is the important "push" factor in migration it is frequently believed to be. Comparisons of the post-migration employment history of the migrants with those of non-migrants at the source-point should help to indicate whether migrants succeeded in improving their employment status by making their move. Similar comparisons with the non-migrants at the destination point can help to indicate whether the migrants are at a disadvantage in competing for economic opportunities with the resident population.

Most studies of these questions have failed to give convincing answers even for a specific time and place. Some studies have failed, because they have focused on "distress" migrants without considering the relative importance of such migrants in the total stream of migration. The result has been a stereotype of the migrant in a role such as that of

* This study was made possible by a grant from the Faculty Research Funds of the Horace Rackham School of Graduate Studies of the University of Michigan.

the "Okie". Other studies have failed to compare the migrants with similar non-migrants at either end of the migration process, so that it is difficult to determine whether the employment history of the migrants was distinctive and whether it was a product of migration or of some other factor.

There is an attempt in the present study to minimize the first difficulty by considering a sample of migrants in the area studied without restriction as to their "distress" position. The second difficulty is minimized by comparing the migrants with carefully selected "control" groups of non-migrants at each end of the migration process. An important methodological objective of this study is to experiment with the use of the matched-group control method in the population-migration field. The authors recognize that the data of this study are limited to a specific time period and to specific streams of migration. Many further studies are needed to permit valid generalizations. It is desirable that such studies should refer to migration between specific areas and should control as many relevant factors as possible.

THE DATA AND METHODS

The data for this study are from the Michigan Population and Unemployment Census of 1935.[1] The distinctive feature of the schedule for this census is a month-by-month history of the occupation and employment status of every person 15 years old or over for the period April, 1930, to January, 1935. This history includes data on place of employment, which makes it possible to identify migrants and their routes of movement.

The sample for the present study consists of all those white male migrants to Flint or Grand Rapids from other places in Michigan, who were at least 25 years old at the time of migration to the cities. Males 15 to 24 years old were excluded from the study in order to deal only with those migrants whose eductaion was presumably completed. The female migrants were reserved for a later study, because their employment status involves special factors. Non-white migrants, originally included in the study were eliminated, because they were found to constitute a negligible proportion of the intra-Michigan migrants to Flint and Grand Rapids. Flint and Grand Rapids were selected as the destination points to be studied, because they are very different both

[1] *Michigan Census of Population and Unemployment*, Michigan State Emergency Welfare Relief Commission: Lansing, 1937, Nos. 1–10. The sampling procedure for this study involved complete enumeration of all cities in Micigan between 3,000 and 40,000 in population, a 20 per cent sample in each city of 40,000 and over, and a complete enumeration of a random selection of 20 per cent of all the towns and villages (under 3,000) and of open country type rural townships.

with respect to population history and economic base. Therefore, it was hoped that studies of the migrants to each place might be treated as separate "experiments." Agreement of the findings for the two cities should give them greater validity. The study was necessarily limited to intra-Michigan migration, because data were not available for non-migrants outside of Michigan. After observing the limitations noted above in selection of the sample, the sample of migrants to Flint was found to number 360. The number in the Grand Rapids sample was 186.

These samples are the basis for a series of studies of the relationship between migration status and a number of other characteristics (education, occupation, occupational mobility). This paper on the relationship of migration and unemployment is only one part of the larger series.

For the study of unemployment, each migrant was "matched" with a "control" non-migrant at the place from which he came and another at the place to which he moved (either Grand Rapids or Flint). The characteristics used for matching were age (within 3 years), occupation (in terms of the major census socio-economic groups), occupational history, (in terms of change between socio-economic classes), education (within 2 years of school achievement), and marital status. For every characteristic, except marital status, the matching was done as of the date of migration. Data on marital status were available only as of the end of the period. Nevertheless it was used as a basis of matching, because it was felt that the error involved would be considerably less than that involved in omitting this factor as a control. In a very few cases migrants in the study had come from places of less than 3,000 population which had not been included in the census sample. Since it was impossible to match such migrants at the actual source-point, they were matched with non-migrants from a similar place located near the actual source-point. Three of the cases of failure to match migrants are accounted for by the failure to find a substitute source-point. The total number of such successfully made substitutions is 9.

Although the choice among "eligible" matches was not made in the most rigorous manner possible, it is believed that the procedure used did not introduce any systematic biases. The procedure was essentially to enter the schedules at a point determined by a system of random numbers, examine the schedules serially until a match was found, then to select a new starting point and to repeat the process. Several alternative methods involving selection of all possible eligible matches and random selection among them were abandoned as prohibitively time-consuming. There does not seem to be any reason to believe that systematic errors were introduced by the method used.

Although it was not possible to match all the migrants at both ends of the migration, proportions matched are large enough so that the authors are confident that the results of the study are not affected by the unmatched cases. Of the 360 migrants to Flint 312, or 87 per cent were matched at the destination and 296, or 82 per cent, at the source-points. Of the 186 migrants to Grand Rapids 171, or 92 per cent were matched at the destination and 149, or 80 per cent at the source-points. This represents a considerably higher proportion of matched cases than are found in most comparable "matched-group" studies. A study of the unmatched migrants indicates that their employment history is not sharply different from that of the matched cases.

Several specific problems in connection with the matching need to be explained. First, those migrants whose occupational classification was "farmer" or "farm-laborer" obviously could not be matched identically for occupation at the destination point. The procedure followed was to match farmers with members of the "proprietors, managers, and officials," group and to match "farm-laborers" with "common laborers." Although these are the groups closest to the farm occupational categories in the census hierarchy of socio-economic classes, the matching on this basis is admittedly makeshift. Therefore, in the specific comparisons of migrants and non-migrants at the destination the effect of including or excluding farm migrants has been evaluated in each case. In any event, the matching permits a comparison of the farm migrants with the urban occupational group with which they are most frequently contrasted and even combined.[2]

In the study of unemployment a second problem arose concerning those persons who were either unemployed or out of the labor force during the entire period from 1930 to date of migration. It was not possible to classify these persons by occupation during the pre-migration period. In each of these cases the match was made as if "unemployment" or "out of the labor force" were the pre-migration occupation, respectively. This means that the employment status of these two categories of migrants was artificially perfectly matched for the pre-migration period. These cases can be compared, however, for the post-migration period for which there were no artificial restraints. During the pre-migration period the comparisons are valid only for those who were employed at some time. In those cases where either a migrant or a control non-migrant was found to be out of the labor force during part of a period, this part of the period was omitted from the comparison

[2] In a number of tabulations of the 1940 U. S. Census for cities farm operators were tabulated with "proprietors, managers, and officials," and farm laborers were tabulated with "laborers."

for the pair. The periods excluded for all groups taken together were relatively small.

We matched 271 of the Flint migrants and 140 of the Grand Rapids migrants at both ends of the migration process. Comparison of tables based on these completely matched (both end) cases with the tables based on migrants matched at only one end of the comparisons have resulted in no essential differences. Therefore, in this paper the tables and discussions are based on the larger number of cases, including one-end matches.

On the average the migrants made their moves about 37 months after the beginning of the period, so the average pre-migration period is about 3 years. One important limitation of the data for this study needs to be mentioned. Data are available only for "resultant" migrants, that is those who moved to Flint or Grand Rapids between 1930 and 1935 and were still there in 1935. Data for the migrants who moved again away from Flint or Grand Rapids before 1935 are not available.

THE FINDINGS

In the pre-migration period, the migrants to Flint or Grand Rapids had a somewhat higher rate of unemployment[3] than the non-migrant control group at the source-points. However, in neither case was the unemployment rate of the migrants or the differential between migrants and non-migrants large enough to justify considering the personal experience of unemployment as the important causal factor in migration.

The data on pre-migration employment status are found in Table 1 for the migrants and their source matches. In comparing migrants and non-migrants during the pre-migration period those unemployed during the entire pre-migration period should be excluded, since these migrants and non-migrants are artificially equated on unemployment. Excluding this group, 18 per cent of the Flint migrants and 13 per cent of the non-migrant source matches were unemployed in the pre-migration period. For Grand Rapids the corresponding figures are 13 per cent and 11 per cent. Even if those totally unemployed during the pre-migration period are included in these rates only 24 per cent of the Flint migrants and 21 per cent of the migrants to Grand Rapids were unemployed in the pre-migration period. The migrants do not appear to be a group marked by an unusually high rate of unemployment for the period in question. In this and other comparisons to follow, differ-

[3] The "unemployment rate" used in this study refers to the percentage of persons in any group who were unemployed at any time during the period in question.

ences will appear as between unemployment rates for Flint and Grand Rapids for each type of migrant or non-migrant. Such differences are to be expected on the basis of the different economic structures of the two cities. However, for the two cities the differences between the migrant and non-migrant unemployment rates are in the same direction in almost every comparison. Variations in the absolute rates of unemployment as between the two cities are not relevant to the problems being investigated.

TABLE 1

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE, FOR PRE-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Pre-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrants	Non-migrants	Migrants	Non-migrants
Total percentage:	100.0	100.0	100.0	100.0
In labor force:	99.0	99.0	99.3	99.3
Always employed	76.0	81.1	85.2	87.2
Always unemployed	5.4	5.4	0.7	0.7
Sometimes unemployed	17.6	12.5	13.4	11.4
Out of labor force	1.0	1.0	0.7	0.7
Number	296	296	149	149

A check on the relative numerical importance of the group of migrants unemployed throughout the pre-migration period was provided by the selection of "random" samples of non-migrants at source and destination points for each group of migrants. The only restriction on selection was that the non-migrants should be at least 25, male, and white. Each migrant group thus had a random non-migrant "comparison" group. These non-migrants were paired at random with migrants to determine a date for separation of the "pre-migration" and "post-migration" periods. At both ends of the migration process and in both cities the number of migrants unemployed throughout the pre-migration period was approximately equal to the number so unemployed among the random comparison groups. Whether considered on an absolute basis or relative to non-migrant comparison groups those unemployed throughout the pre-migration period were not an important group.

It is interesting to not that in the pre-migration period the employment status of the migrants did not compare unfavorably with that of

non-migrant controls at the destination points. As Table 2 indicates, there was less than 1 percentage point difference in the unemployment rates of migrants to Flint and the non-migrant controls in Flint; 19.9 per cent of the migrants and 20.5 per cent of the non-migrants were unemployed in this comparison. The unemployment rate among migrants to Grand Rapids was even less than that of non-migrants in Grand Rapids in the pre-migration period. The unemployment rates were 15.2 and 21.1 for migrants and non-migrants respectively.

TABLE 2

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION FOR PRE-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Pre-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrants	Non-migrants	Migrants	Non-migrants
Total percentage	100.0	100.0	100.0	100.0
In labor force:	98.4	98.4	98.9	98.9
Always employed	71.8	71.2	81.9	76.0
Always unemployed	6.7	6.7	1.8	1.8
Sometimes unemployed	19.9	20.5	15.2	21.1
Out of labor force	1.6	1.6	1.1	1.1
Number	312	312	171	171

The results which have been cited for the pre-migration period were found to be substantially unaltered when separate tabulations were made excluding the migrants with a farm background.

The peak unemployment rate for migrants occurred after rather than before the migration. While the frequency of unemployment among migrants rose sharply in the post-migration period, the unemployment rate for the non-migrant controls showed only a moderate increase at either end of the movement. This is demonstrated by the data in Tables 1–4 to be true whether or not the totally unemployed are included in the pre-migration unemployed group. For example, while the unemployment rate for migrants to Flint increased from 24 to 49 per cent, that for the non-migrant source controls increased from 18 to 24 per cent. While the percentage of increase in unemployment rate was no more than 36 per cent for any non-migrant group, it was no less than 82 per cent for any migrant group. In every case the increase in unemployment rates for the migrant group was significantly

greater[4] than the corresponding increase for its non-migrant control—whether source or destination.

TABLE 3

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION FOR POST-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Post-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrant	Non-migrant	Migrant	Non-migrant
Total percentage	100.0	100.0	100.0	100.0
In labor force:	96.8	97.8	94.1	98.2
Employed	48.4	71.2	57.9	68.4
Unemployed	48.4	26.6	36.2	29.8
Out of labor force	3.2	2.2	5.9	1.8
Number	312	312	171	171

TABLE 4

PERCENTAGE DISTRIBUTION OF EMPLOYMENT STATUS OF MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE FOR POST-MIGRATION PERIOD FLINT AND GRAND RAPIDS

Employment Status in Post-migration Period	Migrant Status and 1935 Residence			
	Flint		Grand Rapids	
	Migrant	Non-migrant	Migrant	Non-migrant
Total percentage	100.0	100.0	100.0	100.0
In labor force:	97.6	98.6	96.0	97.4
Employed	48.6	74.3	58.4	82.6
Unemployed	49.0	24.3	37.6	14.8
Out of labor force	2.4	1.4	4.0	2.6
Number	296	296	149	149

The relatively high post-migration unemployment rate was not restricted to a particular occupational group, although the migrants with a farm background had the highest unemployment rates. Tables 5 and 6 show the unemployment status of each migrant and non-migrant control group separately for 3 major occupational groupings:

[4] The differences were significant at the .01 level in every case but one. The difference for the Grand Rapids migrant and non-migrant destination groups was significant at the .05 level. The differences were significant at these levels whether or not the "totally" unemployed were considered in the pre-migration period.

TABLE 5

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS BY OCCUPATIONAL CLASS, MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE, FLINT AND GRAND RAPIDS, MICHIGAN

Occupational Class before Migration	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed
Total	100.0	50.0	50.0	100.0	75.3	24.7	100.0	64.7	35.3	100.0	85.4	14.6
White collar[2]	100.0	77.0	23.0	100.0	85.3	14.7	100.0	75.4	24.6	100.0	94.8	5.2
Blue collar[3]	100.0	49.0	51.0	100.0	69.5	30.5	100.0	61.4	38.6	100.0	71.1	28.9
Farm workers[4]	100.0	22.5	77.5	100.0	86.4	13.6	100.0	46.4	53.6	100.0	87.1	12.9
Others[5]	100.0	—	—	100.0	—	—	100.0	—	—	100.0	—	—

[1] Exclusive of individuals out of the labor force during entire post-migration period.
[2] Includes: professional workers; proprietors, managers, officials; and clerical workers.
[3] Includes: skilled, semi-skilled, and unskilled workers, and servants.
[4] Includes: farmers and farm laborers.
[5] Includes all persons unemployed or out of labor force during entire pre-migration period. Numbers too small for computation of percentages.

TABLE 6

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS BY OCCUPATIONAL CLASS, MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION, FLINT AND GRAND RAPIDS, MICHIGAN

Occupational Class	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed
Total	100.0	50.0	50.0	100.0	72.8	27.2	100.0	62.4	37.6	100.0	70.6	29.4
White collar[2]	100.0	77.0	23.0	100.0	94.5	5.5	100.0	75.0	25.0	100.0	80.9	19.1
Blue collar[3]	100.0	45.4	54.6	100.0	68.8	31.2	100.0	58.5	41.5	100.0	61.8	38.2
Farm workers[4]	100.0	25.6	74.4	100.0	74.4	25.6	100.0	41.7	58.3	100.0	65.0	35.0
Others[5]	100.0	—	—	100.0	—	—	100.0	—	—	100.0	—	—

[1] Exclusive of individuals not in the labor force.
[2] Includes: professional workers, proprietors, managers, officials, and clerical workers.
[3] Includes: skilled, semi-skilled, and unskilled workers, and servants.
[4] Includes: farmers and farm laborers.
[5] Includes all persons unemployed or out of labor force during entire pre-migration period. Numbers too small for computation of percentages.

TABLE 7

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS, BY EDUCATION, MIGRANTS AND NON-MIGRANTS MATCHED AT SOURCE, FLINT AND GRAND RAPIDS, MICHIGAN

Highest Educational Level Reached	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed
Total	100.0	49.5	50.5	100.0	75.2	24.8	100.0	64.2	35.8	100.0	85.4	14.6
Grade school	100.0	33.3	66.7	100.0	73.5	26.5	100.0	53.3	46.7	100.0	84.4	15.6
High school	100.0	67.3	32.7	100.0	76.0	24.0	100.0	73.7	26.3	100.0	82.0	18.0
College	100.0	79.2	20.8	100.0	83.5	16.5	100.0	89.5	10.5	100.0	95.2	4.8

[1] Exclusive of individuals out of labor force during entire post-migration period.

TABLE 8

PERCENTAGE DISTRIBUTION OF POST-MIGRATION EMPLOYMENT STATUS, BY EDUCATION, MIGRANTS AND NON-MIGRANTS MATCHED AT DESTINATION, FLINT AND GRAND RAPIDS, MICHIGAN

Highest Educational Level Reached	Flint						Grand Rapids					
	Migrants			Non-migrants			Migrants			Non-migrants		
	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed	Total[1]	Em-ployed	Unem-ployed
Total	100.0	50.0	50.0	100.0	72.8	27.2	100.0	64.2	35.8	100.0	72.9	27.1
Grade school	100.0	35.3	64.7	100.0	60.7	39.3	100.0	53.2	46.8	100.0	65.4	34.6
High school	100.0	65.8	34.2	100.0	75.7	24.3	100.0	73.7	26.3	100.0	78.1	21.9
College	100.0	69.2	30.8	100.0	100.0	0.0	100.0	89.5	10.5	100.0	90.5	9.5

[1] Exclusive of individuals out of labor force during entire post-migration period.

white collar workers (professional workers, proprietors and managers, officials, and clerical workers), blue collar workers (skilled, semi-skilled, and unskilled workers, and servants), farm workers (farmers and farm laborers). In each of these occupational categories the post-migration unemployment rate of the migrants to Flint or Grand Rapids was higher than that for comparable non-migrant control groups at the source or destination. It is clear from these data that the farm migrants had a much higher rate of unemployment than any other group whether migrant or non-migrant.

Similarly, the relatively high unemployment rate of the migrants is found at each of three educational levels. In Tables 7 and 8 the employment status of each migrant and non-migrant group is shown separately for those with grade school, high school, and college education, respectively. The only comparison in which the migrants do not have the higher rate of unemployment is that between migrants to Grand Rapids and non-migrants to Grand Rapids for those with some college education. The numbers involved in this comparison are very small.

The sizes of the sub-samples of the occupational groups and educational groups on which the percentages for Tables 5–8 are based are found in Table 9.

TABLE 9

NUMBER IN EACH SAMPLE, BY OCCUPATIONAL CLASS AND EDUCATIONAL LEVEL, FLINT AND GRAND RAPIDS

Occupational Class and Educational Level	Flint		Grand Rapids	
	Sample Matched at Source	Sample Matched at Destination	Sample Matched at Source	Sample Matched at Destination
Occupational class: Total	296	312	149	171
White collar	88	94	60	69
Blue collar	106	113	48	56
Farm	83	79	39	41
Others	19	26	2	5
Educational level: Total	296	312	149	171
Grade school	166	168	86	98
High school	106	117	43	49
College	24	27	20	24

It is recognized that for these various post-migration comparisons it would have been desirable that the matching include control on pre-migration employment status. However, it can be shown that this

matching defect does not seriously affect the findings. In the first place, the difference in *increase* in unemployment rates (not just the absolute post-migration differences in unemployment) as between migrant and non-migrant is significantly greater for migrants in every comparison. Secondly, separate post-migration unemployment rates were computed for those migrants unemployed and those never unemployed in the pre-migration period. In every case both the previously employed and previously unemployed migrant groups had higher post-migration unemployment rates than the non-migrant controls. The difference in unemployment rates as between migrants with a pre-migration history of unemployment and those with none was relatively small.

In this connection it is interesting that the correlation between pre-migration and post-migration unemployment was considerably greater for the non-migrants than for their migrant counterparts. Among the migrants the rate of post-migration unemployment was high both for those who had and those who did not have pre-migration unemployment. On the other hand among the non-migrants the rate of post-migration unemployment was high for those with pre-migration unemployment but very low for those without it. Apparently, for migrants the fact of migration was more important than previous employment status in determining employment status in their new home.

Consideration of the length of unemployment for the unemployed in each group of migrants and non-migrants indicates that the higher unemployment rate of the migrants is not a consequence of short unemployment periods distributed among many persons. Among those unemployed after migration, the average length of unemployment was greater for migrants than for non-migrants in every comparison but one, and in that case the average length of unemployment was the same for each group. This is in part an answer to the idea that the migrant unemployment was of a short-run "frictional" character.

Whatever the effects on the sending and receiving communities, migration appears to have offered no effective solution to unemployment for the migrants, themselves. Quite the contrary, the evidence seems to support the thesis that migration resulted in a deterioration of their employment status. At least for the short-run periods considered in this study, if there is any line of causation between migration and unemployment it appears to run from migration to unemployment rather than the reverse.

Evidence for the thesis that unemployment "causes" migration is sometimes presented in the form of data showing that over a given period of time the unemployment rate was highest for those who

migrated during the period. Thus, it may be said that 60 per cent of the source-matched migrants were unemployed sometime between 1930 and 1935, as compared with only 27 per cent of matched-source non-migrants. Such evidence is really not relevant to the hypothesis, since it fails to indicate whether the unemployment preceded or followed the migration. Although it is true that 60 per cent of the Flint migrants[5] were unemployed at some time during the 5 year period, only 23 per cent of the migrants as compared with 18 per cent of the source non-migrants were unemployed prior to migration. The large differential in unemployment rates is *after* migration, not *before*.

The results of the present study are consistent with the hypothesis that in a depression migrants tend to be at a disadvantage in the new labor market to which they move. This may result from the fact that the migrants lack the specific skills demanded in the local labor market. It may be that, although they have the needed skills, the migrants are at a disadvantage in their poorer knowledge of how to get and hold a job in the local situation, or it may be that in a depression employers are under pressure to favor the local labor supply.

The methodological aspects of this study are equally important with the substantive findings. At least in this instance, the use of matched group control techniques proved to be feasible in a migration study involving a relatively small group. The technique furnishes factor-control frequently lacking in this kind of study. Further experimentation with the technique appears to be warranted.

---

[5] Sixty per cent of the source-matched migrants and 61 per cent of the destination-matched migrants.

MINIMUM X^2 AND MAXIMUM LIKELIHOOD SOLUTION IN TERMS OF A LINEAR TRANSFORM, WITH PARTICULAR REFERENCE TO BIO-ASSAY

JOSEPH BERKSON, M.D.
Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minnesota

An old and commonly used device for fitting a curve $y=F(x, \alpha, \beta)$ is to find a function of y which is linearly related to x, or a function of x which is linearly related to y, or functions of y and x which are linearly related to each other. The linearly related functions are plotted against one another, and a line is fitted to the points, usually by eye, sometimes by "least squares" in terms of the transforms. Texts dealing with empirical curve-fitting characteristically use this scheme [6]. Before Bliss and Fisher [2], no attempt was made to adjust the transforms systematically in order to achieve a fit that fulfilled defined criteria in terms of the original measures y and x, nor was it known that such adjustments were possible. These authors presented for the bio-assay experiment a method in terms of probits, which as shown by Garwood, [5] accomplished a maximum likelihood estimate of the integrated normal curve, when the observations were distributed binomially. Following the procedures used by Bliss and Fisher for the integrated normal curve, I [1] formulated the similar adjustments applicable to "logits" for a maximum likelihood solution of the logistic function, but did not advocate using this solution. Finney [4] has recently presented the adjustments for a maximum likelihood solution of several functions of interest in bio-assay. So far as I know the similar method for a solution fulfilling the criteria of minimum X^2 rather than maximum likelihood has not been presented. In the following paper this is given for several functions in common statistical use, and at the same time a résumé is given of the maximum likelihood adjustments for the same functions.

WE ARE DEALING with a measure Q $(P=1-Q)$ which is a function of x and two parameters α, β, the estimates of these parameters a, b, and Y, a linear transform of Q.

$$Q_i = F(x_i, \alpha, \beta) = F(Y_i) \tag{1}$$

$$Y_i = \alpha + \beta x_i \tag{2}$$

$$\hat{q}_i = F(x_i, a, b) = F(\hat{y}_i) \tag{3}$$

$$\hat{y}_i = a + bx_i. \tag{4}$$

Let

$$z = \frac{\partial \hat{q}}{\partial \hat{y}}; \quad \text{then} \quad \frac{\partial \hat{q}}{\partial a} = \hat{z}, \quad \frac{\partial \hat{q}}{\partial b} = \hat{z}x.$$

If the observation q_i is distributed normally and independently with σ equal at all values of x, then the estimates a and b fulfilling the criteria of either maximum likelihood or minimum X^2 are identical, and are given by the normal equations

$$\sum n_i\hat{z}_i(q_i - \hat{q}_i) = \sum w_i'(q_i - \hat{q}_i) = 0 \tag{5*}$$

$$\sum n_i\hat{z}_i x_i(q_i - \hat{q}_i) = \sum w_i' x_i(q_i - \hat{q}_i) = 0. \tag{6}$$

If q_i is distributed binomially with $\sigma_i^2 = (P_iQ_i)/n_i$, the normal equations for maximum likelihood are given by

$$\begin{aligned} \sum \frac{n_i}{\hat{p}_i\hat{q}_i}(q_i - \hat{q}_i)\frac{\partial \hat{q}_i}{\partial a} &= \sum \frac{n_i}{\hat{p}_i\hat{q}_i}\hat{z}_i(q_i - \hat{q}_i) \\ &= \sum w_i''(q_i - \hat{q}_i) = 0 \end{aligned} \tag{7}$$

$$\begin{aligned} \sum \frac{n_i}{\hat{p}_i\hat{q}_i}(q_i - \hat{q}_i)\frac{\partial \hat{q}_i}{\partial b} &= \sum \frac{n_i}{\hat{p}_i\hat{q}_i}\hat{z}_i x_i(q_i - \hat{q}_i) \\ &= \sum w_i'' x_i(q_i - \hat{q}_i) = 0. \end{aligned} \tag{8}$$

For X^2 we have

$$X^2 = \sum \frac{n_i}{\hat{p}_i\hat{q}_i}(q_i - \hat{q}_i)^2$$

and to minimize we set the first derivatives with respect to a and b equal to zero, yielding

$$\frac{\partial X^2}{\partial a} = -2\sum \frac{n_i}{\hat{p}_i\hat{q}_i}(q_i - \hat{q}_i)\frac{\partial \hat{q}_i}{\partial a}$$

$$\sum \frac{n_i}{(\hat{p}_i\hat{q}_i)^2}(1 - 2\hat{q}_i)(q_i - \hat{q}_i)^2\frac{\partial \hat{q}_i}{\partial a} = 0 \tag{9}$$

$$\begin{aligned} &= \sum \frac{n_i}{(\hat{p}_i\hat{q}_i)^2}\hat{z}_i(\hat{p}_iq_i + p_i\hat{q}_i)(q_i - \hat{q}_i) \\ &= \sum w_i'''(q_i - \hat{q}_i) = 0 \end{aligned} \tag{10}$$

* The primes on the w's are to distinguish between the different weights; they do not represent successive differential coefficients.

and similarly for b

$$\sum \frac{n_i}{(\hat{p}_i\hat{q}_i)^2}\hat{z}_i(\hat{p}_iq_i + p_i\hat{q}_i)x_i(q_i - \hat{q}_i) = \sum w_i'''x_i(q_i - \hat{q}_i) = 0. \qquad (11)$$

The condition for the minimum X^2 (9) differs from the condition (7) of maximum likelihood only in the second term of (9). This term will vanish as (a) $\hat{q}_i \rightleftarrows 0.5$, (b) $q_i \rightleftarrows \hat{q}_i$ so that for experiments such as bio-assay where observations are in the neighborhood of the L.D. 50, or where the observations are close to the curve the two solutions may be expected not to differ very much.[1]

It is seen that all the normal equations are of the form

$$\sum w(q - \hat{q}) = 0$$

or

$$\sum wq = \sum w\hat{q}.$$

It is interesting to note therefore that both minimum X^2 and maximum likelihood, for normal or binomial variation, impose the simple requirement that the weighted average of the estimates be equal to the weighted average of the observations, the only difference among them being the value of the weights w.

For the situations usually considered in bio-assay work, the normal equations cannot be solved directly because the functions usually employed are not linear in the parameters, and where q is assumed to vary binomially there is the additional reason that the weights are a function of the $\hat{q}$'s, the values to be estimated. The equations may be solved by a procedure of successive approximations by making the following substitutions. Let $\hat{q}_{i,1}$ be a preliminary estimate of $\hat{q}_i$, $\hat{p}_{i,1}$ the corresponding value of $\hat{p}_i$ and $\hat{z}_{i,1}$ the corresponding value of $\hat{z}_i$, and for brevity write $\hat{q}_1$ for $\hat{q}_{i,1}$, $\hat{p}_1$ for $\hat{p}_{i,1}$ and $\hat{z}_1$ for $\hat{z}_{i,1}$.

$$q_i - \hat{q}_i \equiv (q_i - \hat{q}_1) + (\hat{q}_1 - \hat{q}_i).$$

$(\hat{q}_1 - \hat{q}_i)$ is replaced by the first term in a Taylor's expansion

[1] It is sometimes stated that it has been proved that the minimum X^2 and maximum likelihood solutions converge to the same solution as $n \rightleftarrows \infty$. From the foregoing it seems to me that this is not necessarily so. It will be so if $q \rightleftarrows \hat{q}$ as $n \rightleftarrows \infty$, but this is not always the case. When the function we are fitting is the "true" curve, theoretically $(q - \hat{q}) \rightleftarrows 0$ with probability approaching 1 as $n \rightleftarrows \infty$, but not if we are fitting a function that is different from the "true" function, as is usually the case in statistical curve fitting. The same is true if we are fitting a line by least squares minimizing the squared residual of y, to observations of a random sample from a bivariate normal distribution, when there is an error in the observation of x.

TABLE 1
WORKING VALUES AND WEIGHTS

Function	Transform, $\hat{y}$	$\hat{z}=\frac{\partial q}{\partial \hat{y}}$	Working value	Weight, w		
				Normal	Binomial	
				Max. lik.; min. X^2	Max. Lik.	Min. X^2
Integrated normal curve $\hat{q}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t=a+bx} e^{-t^2/2}$	Normal deviate, $\hat{t}$ or Probit $=\hat{t}+5$	$\hat{z}$, ordinate of the normal curve at $\hat{t}$ $\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$	$\hat{t}_1+\frac{q-\hat{q}_1}{\hat{z}_1}$	$n\hat{z}_1^2$	$\frac{n\hat{z}_1^2}{\hat{p}_1\hat{q}_1}$	$\frac{n\hat{z}_1^2}{(\hat{p}_1\hat{q}_1)^2}\left(\hat{p}_1q+p\hat{q}_1\right)$
Logistic $\hat{q}=\frac{1}{1+e^{-(a+bx)}}$	Logit* $\hat{l}=\ln\frac{\hat{q}}{\hat{p}}$	$\hat{p}\hat{q}$	$\hat{l}_1+\frac{q-\hat{q}_1}{\hat{p}_1\hat{q}_1}$	$n(\hat{p}_1\hat{q}_1)^2$	$n\hat{p}_1\hat{q}_1$	$n(\hat{p}_1q+p\hat{q}_1)$
Sine $\hat{q}=\sin^2(a+bx)$	Unchristened $\hat{k}=\sin^{-1}\sqrt{\hat{q}}$	$2\sqrt{\hat{p}\hat{q}}$	$\hat{k}_1+\frac{q-\hat{q}_1}{2\sqrt{\hat{p}_1\hat{q}_1}}$	$n\hat{p}_1\hat{q}_1$	n	$\frac{n(\hat{p}_1q+p\hat{q}_1)}{\hat{p}_1\hat{q}_1}$
Exponential $\hat{q}=e^{a+bx}$	Natural logarithm $\hat{h}=\ln\hat{q}$	$\hat{q}$	$\hat{h}_1+\frac{q-\hat{q}_1}{\hat{q}_1}$	$n\widehat{\ }_1^2$	$\frac{n\hat{q}_1}{\hat{p}_1}$	$\frac{n(\hat{p}_1q+p\hat{q}_1)}{\hat{p}_1^2}$

* The sign of the "logit" as given here is the negative of that as defined previously [1].

$$
\begin{aligned}
(\hat{q}_1 - \hat{q}_i) &\doteq \hat{z}_1(\mathfrak{y}_1 - \mathfrak{y}_i) \\
(q_i - \hat{q}_i) &\doteq (q_i - \hat{q}_1) + \hat{z}_1(\mathfrak{y}_1 - \mathfrak{y}_i) \\
&\doteq \hat{z}_1\left[\left(\frac{q_i - \hat{q}_1}{\hat{z}_1} + \mathfrak{y}_1\right) - \hat{y}_i\right] \\
&\doteq \hat{z}_1(\mathfrak{y}_i' - \mathfrak{y}_i).
\end{aligned}
\tag{12}
$$

The value of $\mathfrak{y}_i' = (q_i - \hat{q}_1)/(\hat{z}_1) + \mathfrak{y}_1$ is called the "working value." If we replace $(q_i - \hat{q}_i)$ as in (12) in any of the normal equations previously given, and, where in any of those equations $\hat{z}_i$ appears, substitute the preliminary approximation $\hat{z}_1$, since $\mathfrak{y}$ is linear in a and b, the equations may be solved by the usual procedures as used for fitting a

TABLE 2

FIT OF LOGISTIC FUNCTION*

Iteration	Minimum X^2				Maximum likelihood			
	a	b	X^2	$L+30$†	a	b	X^2	$L+30$†
Original‡	−2.972	2.515	3.9480	0.7988	−2.972	2.515	3.9480	0.7988
1	−3.156	2.473	2.8404	1.0278	−3.290	2.581	2.9943	1.0644
2	−3.140	2.467	2.8390	1.0242	−3.324	2.602	3.0658	1.0657
3	−3.143	2.469	2.8389	1.0255	−3.325	2.602	3.0666	1.0657
4	−3.143	2.469	2.8389	1.0252	−3.325	2.602	3.0666	1.0657
Approximate minimum X^2, noniterative method								
	−3.126	2.444	2.8501	1.0074				

* Example from Finney [3]. To simplify the calculations x, the log dose, as given by Finney was adjusted to make x = zero for the lowest dose, and since the doses proceeded in multiples of 2, it was divided by log 2. The iterations were continued until the absolute difference between successive values of both a and b was less than 0.0005.

† $L = \Sigma r \log \hat{p}_i + \Sigma(n_i - r_i) \log \hat{q}_i$, where n_i is the number exposed at dose x_i, p_i is the fraction affected at x_i, $r_i = p_i n_i$ and $\hat{p}_i$ is the solution value of $p_i(\hat{q}_i = 1 - \hat{p}_i)$.

‡ Solution using n alone as weight.

straight line by least squares, using as weight $w_i = w_i'\hat{z}_1$, $w_i = w_i''\hat{z}_1$ or $w_i = w_i'''\hat{z}_1$ as the case may be. With a solution obtained in this manner, the procedure outlined is repeated, using for the evaluation of the new weights and working values, the values of $\hat{q}$ obtained in the present solution, and by repeating this, the respective minimum X^2 or maximum likelihood solution is approached as a limit.

In Table 1 are given for some functions commonly used in statistical practice the working values and weights in the case of normal and bi-

nomial variation, for the minimum X^2 and maximum likelihood solutions. In Table 2 are given the results of fitting a logistic function to some data used by Finney [3] by the methods outlined for (1) minimum X^2; (2) maximum likelihood; (3) noniterative approximate minimum X^2 solution previously suggested by me [1].

REFERENCES

[1] Berkson, J., "Application of the Logistic Function to Bio-Assay" (1944). *Journal of the American Statistical Association*, 39, 357.

[2] Bliss, C. I., with an Appendix by Fisher, R. A., "The Calculation of the Dosage Mortality Curve" (1935). *Annals of Applied Biology*, 22, 134.

[3] Finney, D. J., "Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve." Cambridge University Press, 1947, p. 201.

[4] Finney, D. J., "The Principles of Biological Assay" (1947). *Journal of the Royal Statistical Society*, (*Suppl.*) 9, 46.

[5] Garwood, F., "The Application of Maximum Likelihood to Dosage Mortality Curves" (1940). Biometrika, 32, 46.

[6] Lipka, Joseph, "Graphical and Mechanical Computation." Ed. 1, New York, John Wiley & Sons, Inc., 1918, 264 pp.

SOME INADEQUACIES OF THE FEDERAL CENSUSES OF AGRICULTURE*

RAYMOND J. JESSEN
Bureau of Agricultural Economics, and Statistical Laboratory, Iowa State College

INTRODUCTION

A NUMBER OF criticisms of the federal censuses of agriculture have been made by users of census data. The primary purpose of this paper is to bring together for discussion some of these criticisms, and to propose modifications designed to make the censuses more adequate and useful (i) as sources of general statistics on agriculture and (ii) as sources of information for economic analyses. A secondary purpose is to consider some questions of survey methods for obtaining agricultural information—whether the survey be a census or a sample. The results of a recent statistical survey of the United States[1] are used as an aid in determining the general order of magnitude of some of the points under consideration.

The definition of a farm used by the Bureau of the Census appears to be inadequate for the purposes of agricultural economists on two counts: the exclusion of certain agricultural producing units because of (i) the *scale of operations* and (ii) the *type of enterprise*. In this connection it is estimated from the survey that the 1945 Census of Agriculture excluded, by the definition of a farm adopted, a group of about 5,200,000 units whose scale of agricultural operations was too small to qualify them as "census farms" but which contributed an estimated $445,000,000, or about 1.8%, to total agricultural production in 1946. This in itself may not be important, but where *kinds* of agricultural production are under examination it may be of practical significance.

In addition to the omissions from the census of units that do not meet the definition of a farm, there is some evidence that the census may have failed to enumerate some farms that do meet the definition. This is true particularly for comparatively small farming operations. The apparent deficiencies in the coverage of farms by the census may have arisen because of an incompleteness of coverage by the census inves-

* Journal paper No. J-1635 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 1005, in cooperation with the Bureau of Agricultural Economics, U. S. Department of Agriculture. The views and conclusions expressed in this paper are entirely those of the author and do not necessarily in any way reflect the views of the Bureau of Agricultural Economics.

[1] A national survey of the household refrigeration market conducted by the Statistical Laboratory in January, 1947. This survey was not designed for this purpose but it does have some information which may be of interest here.

tigator, or because of differences in the interpretation of the value of products of certain operations by the census investigator and by the investigators in the refrigeration survey.

In connection with the problem of choice of area smaller than the county for presentation of data, it is suggested that a "natural" area such as a valley or some such drainage area be adopted by the Bureau of the Census, rather than the present unit which is mainly political and indeed subject to change from time to time. This would provide statistics on a more useful and permanent basis almost throughout the U.S., and particularly in the west where political divisions are of especially small agricultural utility.

General discussion of the problem. The data from federal government censuses are used by a great many persons and for a great many purposes. For the present discussion criticism will be considered mainly from the standpoint of its use for research in agricultural economics. Workers in this field frequently make use of the total farm count, area under various crops, inventories of a number of farm quantities, characteristics of the farm dwelling, etc. Agricultural statisticians, such as those of agricultural Estimates, Bureau of Agricultural Economics, use the data as a "bench mark" on which to base a large body of current estimates on acreages under different crops, crop yields, livestock numbers and production, etc.

Three classes of users of census data in agricultural economics will be distinguished: (1) those who are interested in the number and characteristics of "farms" as firms, (2) those who are interested in the number and characteristics of "farmers" and their families and (3) those who are interested in statistical aggregates—the total production of rice, for example. (There are of course other classes such as those interested in land owners, for example, but only these three classes will be dealt with here.) The first group requires data to be given in the form of frequency tables in which the farm is the unit (examples: Number of farms by size in acres, and by value of output). The second group requires the data also in frequency tables but with the farmer (operator) as the unit (example: income of farm operators by age group). The third group is generally not concerned with the unit by which an item is collected or tabulated but merely the unit of measurement (bushel, head, etc.). It is possible, however, that there is a demand on the part of this group for such frequency distributions as acreage of harvested corn by size of field, yield of apples by age of tree, etc.

The conflict between *farm* and *farmer* as a unit of enumeration and

tabulation (class 1 versus class 2) is of real importance because many farms are "operated" by more than one farmer—for example, partnerships.

In attempting to satisfy the demands of all classes of users, it was of course necessary for the Bureau of the Census to make some compromises which, in part at least, gave rise to the inadequacies to be discussed here. Consideration will be given to those inadequacies arising from:

(1) definition of farm
(2) enumeration
(3) definition of farmer
(4) definition of agriculture
(5) unit for presentation of data for small areas.

Inadequate definition of farm. In order to discuss properly the implications of the "farm" as regarded by the Bureau of the Census, it may be appropriate to quote in full the definition given the enumerators in the 1945 census of agriculture. This definition follows.

> A farm, for Census purposes, is all the land on which some agricultural operations are performed by one person, either by his own labor alone or with the assistance of members of his household, or hired employees. The land operated by a partnership is likewise considered a farm. A "farm" may consist of a single tract of land, or a number of separate tracts, and the several tracts may be held under different tenures, as when one tract is owned by the farmer and another tract is rented by him. When a landowner has one or more tenants, renters, croppers, or managers, the land operated by each is considered a farm. Thus, on a plantation the land operated by each cropper, renter, or tenant should be reported as a separate farm, and the land operated by the owner or manager by means of wage hands should likewise be reported as a separate farm.
>
> Include dry-lot or barn dairies, nurseries, greenhouses, hatcheries, fur farms, mushroom cellars, apiaries, cranberry bogs, etc.
>
> Do not include "fish farms", "fish hatcheries", "oyster farms", and "frog farms". *Do not report as a farm any tract of land less than 3 acres, unless its agricultural products in 1944 were valued at $250 or more.*
>
> Farming, or agricultural operations, consists of the production of crops or plants, vines, and trees (excluding forestry operations) or of the keeping, grazing, or feeding of livestock for animal products (including forestry operations) or of the keeping, grazing, or feeding of livestock for animal products (including serums), animal increase, or value increase. Livestock, as here used, includes poultry of all kinds, rabbits, bees, and fur-bearing animals in captivity—in addition to mules, asses, burrows, horses, cattle, sheep, goats, and hogs. Frequently certain operations are not generally recognized as farming. This is especially true where no crops are grown or where the establishments are not commonly considered as farms.

Following is a partial list of types of specialized agriculture and of operations not generally recognized as farms or farming, for which returns on the Farm and Ranch Schedule are required, provided the area is 3 acres or more or, if less than 3 acres, the value of the products in 1944 was $250 or more: Apiaries (bee farms), Community or cooperative gardens, Country estates and country homes (if there is production of vegetables, eggs, milk, or other agricultural products either for home use or for sale), Cranberry bogs, Dry-lot or barn dairies, Feed lots, Fur farms (fox, mink, skunk, etc., in captivity), Garbage-feeding hog yards, Greenhouses, Hatcheries (baby chicks, poults, etc.), Institutional farms (connected with schools, prisons, hospitals, etc.) Mushroom cellars, Nurseries (except for reforestation projects, or in connection with parks), Part-time farms (agricultural operations incidental to other occupation), Grazing or pasturing of livestock, Harvesting of grass seed, Keeping of chickens and the production of broilers (including battery-laying and battery-broiler plants), Production of medicinal or drug plants and herbs, Production of flowers and bulbs for sale, Production of vegetables under glass, Production of vegetable and flower seeds, plants, bulbs, tubers, etc., Production in captivity of pheasants, quail, etc., Production of mint, spices, or other special crops, Raising of Shetland or other ponies, Rabbit raising, Squab raising.

If any specialized or unusual types of agriculture such as those mentioned above are reported, list type under Supplemental Information on page 12.

Although columns are not provided on the schedule for obtaining reports for all the above-mentioned specialized operations in detail, be sure to report on all items that are applicable, making use of inquiries, for "other crops". Note that value of land and buildings and value of sale of products should be reported in all cases.

Include in one report all such land which the operator uses for agricultural purposes, as previously defined, also all outlying or separate field, meadows, pastures, woodland, and waste land. A farm may consist of two or more separate tracts not necessarily adjacent. Do not include public or open range neither owned nor leased by the operator. If the operator cuts hay from land that he does not own and for which he pays no rent, include such acreage under Wild Hay Cut and explain under supplemental information. Large areas of land or other non-agricultural land held as a separate business and not used for pasture or grazing should not be included.

The following types of establishments and operations do not require returns on the Farm and Rranch schedules unless there are also agricultural operations: Canneries, Cheese factories, Creameries, Deer parks, Fish, frog, alligator, or snake "farms", Fish hatcheries, Game preserves, Kennels, Livestock dealers (except feed lots or other farming operations), Ostrich "farms", Oyster "farms" Parks, Riding academies with no farming operations, Shipping pens, Turpentine "farms" or turpentine "orchards", Distilleries, gins, dryers, mills, refineries, or packing plants, Establishments of 3 acres or more, even though locally known as "farms" on which there are no agricultural operations. Idle or abandoned farms which were not operated in 1944 and will not be operated in 1945, Cutting or gathering of forest products with no farming operations, Landscaping, or maintaining grounds,

and growing of flowers, shrubs, and ornamental etc., except where the land is maintained primarily for their production. Production of maple sirup or sugar with no farming operations, raising canaries, guinea pigs, white rats, or white mice, stock yards and auction yards or barns, trapping of wild animals.

It will be observed that the definition given above excludes certain agricultural producing units because of (i) scale of operations (as measured in acres and in dollars) and other elements are excluded because of (ii) *type* of *enterprise*. Later (Section 6) the effects of type of enterprise will be dealt with when the question of the scope of agriculture is being considered.

An estimate of the *amount* of agricultural production excluded by this definition because of insufficient scale is provided by a recent national sample survey.[3] According to this survey the total agricultural production (stated by producers)[4] was about $25 billion in 1946 (table 1). Of this quantity, $445 million, or 1.8% took place on "subcensus" farms (those too small to qualify as census farms). In general, units large enough to qualify as census farms and whose "operators" reside in the open country, produce crops and animals in about the same dollar volume, but all smaller units whether operated in the cities, towns or villages produce crops valued at about 5 times that of animals (Table 2).

[3] The National Refrigeration Survey dealt with a cross-sectional sample of households both "farm" and "non-farm". A special attempt was made to get households to report agricultural production however small, by avoiding questions referred to *farm* or *ranch* when the respondent said he had no *farm*. In these cases the interviewer was instructed to tell his respondent that even though he did not have a *farm* he might nevertheless be engaged in the production of some crops and animals on his *place* and proceeded with the interview asking for details on such activities. In the office these data were used to classify those units which qualified, as "census farms," the smaller producers as "subcensus farms". A weakness in the survey was the inadequate means for detecting multi-operator farms and therefore their inclusion in the sample.

[4] Includes production used at home. It is realized, of course, that this information reported by the respondent, is subject to error due to vagueness in concept, memory, etc.

The question, put to all households, whether they appeared to engaged in agricultural operations or not, were:

"Does head of household operate a farm?"

(if yes) "How much would you say the garden and field crops sold last year were worth at local market price? (Include home consumed as well as sold.)
How much would you say the animals sold or the products obtained from them and sold last year (such as milk, eggs, furs, honey) were worth at local market price? (Include home consumed as well as sold.)

(if no) "Even though you don't have a farm or regard yourself as a farmer, do you have a garden, chickens or some other small enterprise for production of agriculture products?
(If, No) no further agricultural questions.
(If, Yes) "How much were garden vegetables, fruits, berries etc. worth at the local market price? (Include home consumed as well as sold.)
"How much would you say that animals or animal products were worth at local market price? (Include home consumed as well as sold.)
(Also questions on acreage, etc.)

TABLE 1

ESTIMATED NUMBER AND VALUE OF AGRICULTURAL PRODUCTION IN 1946, OF FARMS AND "SUBFARMS", U. S. BY ZONE

(1)	(2)	(3)	(4)
Zone[a]	Estimated "census" & "subcensus" farms (that is all units reporting some agricultural production)[b]	Estimated "census" farms (those which qualify as farms according to census definition) and relative standard error[c]	Estimated "subcensus" farms (those excluded from the census because scale too small) and relative standard error (2)-(3)
Total			
Number	11,788,000	6,582,000± 5%	5,207,000±12%
Production	$24,957,000,000	$24,512,000,000±10%	$445,000,000±12%
Open Country			
Number	5,946,000	5,200,000± 5%	746,000±18%
Production	$19,784,000,000	$19,706,000,000±11%	$ 77,000,000±18%
Rural Place			
Number	2,857,000	890,000±13%	1,966,000±17%
Production	$ 2,972,000,000	$ 2,787,000,000±28%	$185,000,000±17%
Urban Place			
Number	2,985,000	491,000±39%	2,494,000±22%
Production	$ 2,202,000,000	$ 2,018,000,000±46%	$184,000,000±22%

[a] Zone refers to where the operator lives—not necessarily where farm is situated.
[b] Estimates from National Refrigeration Survey.
[c] Probability is about ⅔ that the difference between the sample estimate and the *true value* (the value that would result from a 100% sample) will be within plus or minus *one standard error.*

It is estimated that there are about 5,207,000 subcensus units producing some agricultural products (this number is not very meaningful because of the obvious indefiniteness of "*no* production"). Nearly half, or 2,494,000 of the operators of these subcensus farms reside in urban towns and cities; another 1,966,000 or 40% of the operators live in rural towns and villages and the remaining 746,000 or 14% are in the open country.[5] The place of residence of the operator may not be the same as the location of the unit. No detail was obtained in the survey to indicate the kinds of agricultural activity in which the operators of these units are engaged, but considering their location and scale of operations it would be reasonable to conjecture that they produce

[5] The partitioning of the U. S. into three "zones" follows a *practice* made possible by and established in connection with the Master Sample of Agriculture. By "urban" zone is meant the area covered by cities and towns having a population, in the 1940 Census of Population of 2500 or more, or otherwise designated by the Bureau of the Census as "urban". The "rural place" zone consists of all named places having at least 100 inhabitants, and other areas with a population density of at least 100 persons per square mile, which are not included in the urban group. These places, whether incorporated or not have had boundaries described around them as part of the Master Sample Project operations. The "open country" zone includes the remaining area of the U. S. outside of the "urban" and "rural place" zones.

mainly poultry and eggs, milk, vegetables, fruits, nuts, rabbits and the like. These are items that may be of importance in food and diet statistics.

Not only does the census definition of farm exclude a segment of agriculture on the basis of scale of operations, but it does so without consistency through time. The total number of farms enumerated by the census, other things being equal, depends on the agricultural price level. During years of high agricultural prices a greater number of units should qualify as farms than during years of low agricultural prices—even though physical production remains the same. This is due to the fact that one of the criteria chosen to measure scale is agricultural production in terms of dollar value ($250 per year). This sensitivity of the census farm count to the price situation could give rise to unfortunate complexity in the interpretation of a time series of farm counts. Specu-

TABLE 2

VALUE OF CROP PRODUCTION AND LIVESTOCK PRODUCTION 1946, OF "SUBCENSUS" AND "CENSUS" FARMS, U.S. BY ZONE[a]

(1) Zone	(2) Source		(3) Total Census and Subcensus Farms	(4) Census Farms and Relative Standard Error	(5) Subcensus Farms and Relative Standard Error
Total, U. S.	Total	No.	11,788,000	6,581,000± 5.4	5,207,000±12.3
		Value	24,957,000,000	24,512,000,000± 9.9	445,000,000±12.1
	Crops	No.	11,225,000	6,217,000± 5.5	5,008,000±11.8
		Value	14,267,000,000	13,894,000,000±14.5	373,000,000±12.4
	Animals	No.	6,891,000	5,553,000± 6.1	1,338,000±24.4
		Value	10,690,000,000	10,618,000,000±10.6	72,000,000±24.3
Open Country	Total	No.	5,946,000	5,200,000± 5.3	746,000±17.7
		Value	19,783,000,000	19,706,000,000±10.8	77,000,000±18.1
	Crops	No.	5,587,000	4,913,000± 5.8	674,000±19.2
		Value	10,252,000,000	10,191,000,000±16.1	61,000,000±19.2
	Animals	No.	4,901,000	4,663,000± 6.2	238,000±40.9
		Value	9,531,000,000	9,515,000,000±11.3	16,000,000±36.4
Rural Place	Total	No.	2,857,000	890,000±12.7	1,967,000±16.6
		Value	2,972,000,000	2,787,000,000±27.6	185,000,000±17.1
	Crops	No.	2,783,000	843,000±13.1	1,940,000±16.9
		Value	2,110,000,000	1,955,000,000±35.0	155,000,000±18.4
	Animals	No.	1,219,000	620,000±14.6	599,000±29.7
		Value	861,000,000	832,000,000±32.5	29,000,000±26.5
Urban Place	Total	No.	2,985,000	491,000±39.2	2,494,000±22.3
		Value	2,202,000,000	2,019,000,000±46.1	183,000,000±22.4
	Crops	No.	2,855,000	461,000±38.4	2,394,000±22.0
		Value	1,905,000,000	1,748,000,000±54.5	158,000,000±22.0
	Animals	No.	771,000	270,000±53.9	501,000±50.8
		Value	298,000,000	271,000,000±64.7	27,000,000±52.4

[a] Estimated from the National Refrigeration Survey.

TABLE 3

ESTIMATED NUMBER AND VALUE OF AGRICULTURAL PRODUCTION 1947, OF CENSUS AND SUBCENSUS FARMS, U.S. BY SIZE (IN ACRES) CLASS[a]

(Numbers in thousands, value in $ millions)

(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)
Value Class	Total		0–3	3–29	30–99	100–179	180–259	260–499	500 & over
1– 99	No.	3,299	3,046	151	75	21			6
	Value	137	127	7	3	*			*
100– 249	No.	2,654	2,160	342	108	38	3	3	
	Value	401	318	57	18	7	*	1	
250– 399	No.	690	303	257	112	15	3		
	Value	221	93	85	38	4	1		
400– 499	No.	437	164	117	111	33	3	9	
	Value	182	67	48	46	15	2	4	
500– 999	No.	1,014	194	330	308	137	39	6	
	Value	693	127	223	211	104	24	4	
1,000– 1,999	No.	912	15	237	414	145	54	33	14
	Value	1,253	20	309	565	213	81	45	20
2,000– 4,999	No.	1,354	14	86	388	464	190	143	69
	Value	4,248	33	254	1,153	1,457	586	511	254
5,000– 9,999	No.	738	—	30	74	193	138	198	105
	Value	5,253	—	219	488	1,319	1,008	1,439	780
10,000–19,999	No.	531	—	32	50	78	71	127	173
	Value	6,968	—	383	660	1,048	960	1,736	2,181
20,000 & over	No.	159	—	33	18	25	11	27	45
	Value	5,601	—	1,293	1,020	589	355	746	1,598
Total	No.	11,788	5,896	1,615	1,658	1,149	512	546	412
	Value	24,957	785	2,878	4,202	4,756	3,017	4,486	4,833

[a] Data from National Refrigeration Survey based on a sample therefore subject to sampling variation. See tables 1 and 5 for indications of the size of these sampling errors.

* Less than $500,000.

lating with data available from the survey (Table 3) it appears that if the agricultural price level is doubled (that is roughly what took place between the 1940 and 1945 census) about 1,500,000 farms should appear in the 1945 Census which would not have qualified for the 1940 census, assuming of course, the number and physical volume of production of all units producing agricultural products remains the same. Also if the price level should drop to one-half the 1945 level about 500,000 farms would be "lost". These speculations assume of course that enumerations each time are complete and that the definition of a farm

used in 1945 was strictly adhered to. These assumptions are not completely true, as will be pointed out in section 4. This sensitivity could be largely eliminated if the volume of production requirement in terms of dollar value were put at a value level each time of census such that a constant physical volume requirement would be maintained. Also if farm counts were published separately by zones, such as the three used here (open country, urban place and rural place), it might be possible to confine most of the changes due to price level fluctuations to the tables for towns and cities—the open country being somewhat freer from such effects.

Inadequate enumeration. Another part of agriculture that has been omitted by the Bureau of the Census, but included in its definition, is that which has been missed for one reason or another by the field worker, deleted in editing, or missed for other reasons. The differences shown in Table 4 between the survey estimates for 1947 and the data

TABLE 4

NUMBER OF FARMS BY CRITERION OF CENSUS DEFINITION SATISFIED, NATIONAL REFRIGERATION SURVEY (1947) AND 1945 CENSUS OF AGRICULTURE

Class[1]	National Refrigeration Survey, Jan. 1, 1947 (estimate and standard error)	Federal Census of Agriculture April 1, 1945
Total "Census" Farms, U S	6,582,000± 5%	5,859,000
Qualifying on:		
Value alone	690,000±20%	99,000
Acreage alone	747,000±20%	552,000
Both acreage and value	5,145,000± 6%	5,208,000

of the 1945 Census of Agriculture are within the range given by twice the sampling error of the survey estimates, for both the total number of farms and value of agricultural production (when adjustment is made for the difference in price level between 1944 and 1946). However, if the number of farms is considered by classes according to the criterion (acreage, value or both) of the Census definition satisfied, it appears that about 590,000 of the total difference of 723,000 between the National Refrigeration Survey and the 1945 Census of Agriculture occurs in the class of farms qualifying on value of products alone (Table 4). The difference for this class is statistically significant.

There are a number of possible explanations of this difference. A reasonable interpretation is that farms of this type are often missed in the census. It appears that these farms are mostly small in scale and, consequently, might not have looked like farms to the Census investigator.

As a rule they were situated in areas generally regarded as "non-farm" or had no resident operator and therefore might more easily have been passed over (Table 5). On the other hand, it might be that they were recognized by, or known to, the enumerator, but in his estimation were so costly to seek out and interview that he decided to omit them.

It is reasonable to believe that at least part of the discrepancy between the NR Survey and the census may be due to differences in the elicitation of value of products.[6] That is, a census investigator might well have visited a unit and found that the value of products (as obtained by him) was not sufficient to qualify the unit as a census farm,

TABLE 5

ESTIMATED NUMBER OF FARMS BY ZONE: 1940 CENSUS OF AGRICULTURE AND RESULTS OF NATIONAL REFRIGERATION SURVEY*

(1)	(2)	(3)
Zone	1940 Census of Agriculture.[a]	Survey estimate[b] and its standard error[c]
Total	6,096,799	6,582,000 ± 5.4%
Open country	5,532,374	5,200,000 ± 5.5%
Rural places	491,062	891,000 ± 13.6%
Urban places	73,363	491,000 ± 38.3%

* The two sets of estimates (1940 and 1947) are not directly comparable because of the fact that the estimates for the 1940 Census of Agriculture are based on location of the farm, while those from the National Refrigeration Survey are based on residence of the farm operator. These latter estimates (1947) tend to indicate more farms in the "urban places" and "rural places" zones. Thus, in the 1945 Census of Agriculture 340,000 farm operators reported themselves as not living on the farm they operated. Although place of residence was not tabulated for these operators, most of them undoubtedly resided in urban or rural places, rather than elsewhere in the open country zone.

[a] These estimates were made in connection with the Master Sample Project before the 1945 data were available. None have been made on the more recent data. Since they are estimates they are subject to error. Zone refers to location of farms and not residence of farm operator as in the Survey Estimates.

[b] Estimated from the National Refrigeration Survey sample. The sampling error of the estimate is given along side each figure. Zone refers to where the operator lives—not necessarily to where farm is situated, as in the estimates of 1940.

[c] Probability is about 2/3 that the difference between the sample estimate and the true value (the value that would result from a 100% sample) will be within *plus* or *minus one standard error.*

while for the same unit a survey investigator might have obtained a reported value of products that would satisfy the census criterion. In the absence of information as to how well value of products was elicited in the census and in the survey, and partly also because of the changes in price level, it is not possible to say which of the two figures for number of farms is more nearly correct. That is immaterial, however, for

[6] It should be noted that the procedure and questions for arriving at the total value of farm products were different in the National Refrigeration Survey than in the 1945 Census of Agriculture.

this analysis. The striking feature is that it is possible for two sets of presumably qualified investigators to take the same definition out in the field, and bring back significantly different results.

"Correction of this inadequacy will require more than just bringing pressure to bear on the workers for better work. Assuming a change in definition, changes in survey procedure may still be recommended." First, out of consideration of economy, it may be feasible to confine the canvassing of cities and towns to a sample but retain the complete coverage of the open county zone (Table 5). Second, require the investigator to call on *all households* in his district (except those in towns and cities which are not in his sample) and make out a report for each—whether the household operates a "farm" or not. The identification of farms would be made in the central office where a consistent policy could be established and kept under control. This procedure is proposed to eliminate or reduce unknown incompleteness in return for one which has a calculable amount of sampling fluctuation. The complete census of all households in the open country whether farm or non-farm would probably have considerable use to market researchers, students of population problems, etc., who now must depend on data obtained on a somewhat unsteady and not too well-known base.

Inadequate definition of operation (farmer). In general the word *operator* (when dealing with farms) is used to describe *a person* who "operates" a farm. If two persons jointly operate a farm both would generally be regarded as operators and each partner would so regard himself. Complete independence in the operation of a farm is therefore not required before the status of "operator" can be reached. It may be helpful to define some concepts on this matter as follows: Let the entrepreneurial function required of each farm be performed by a person or a number of persons which collectively will be called the *operatorship.* Each farm must have one and only one operatorship but that operatorship may consist of one or more persons called operators. Now an operator may be associated with more than one farm through membership in more than one operatorship so the total number of operators for a given area could be a different figure than the total number of farms for that area.

A distinction of this sort is not made by the Bureau of the Census. To quote its instructions, "For a farm operated by two or more partners enter *only one of the partners* as the operator, preferably the senior partner, unless the junior partner is actually conducting the operations." (Italics mine.) This definition of operator, for the sake of simplicity, does violence to reality.

Forcing multi-person operatorships into this mold brings about some curious results. There will be an undercount of the number of true operators—that is, the kind which the economists, sociologists and the laymen would be thinking of. Likewise a number of other undercounts result such as number of farm dwellings having electricity, etc. (only one dwelling from each farm is eligible for this information), mortgage debt (asked of the one operator only and of only those who own land), land owned by the operator personally (here again asked only of the one operator), etc. This difficulty could be overcome by adopting a more complex model than now used—one permitting of multi-person operatorships.

Inadequate definition of agriculture. Although the Bureau of the Census calls its quinquennial survey a "census of agriculture" there is some doubt that it is comprehensive. This "census of agriculture," for example "covers" about 60% of the land area of the United States (Table 6). There may be some objection to the regarding of the remaining 40% as containing no "agriculture". For example, land under forest

TABLE 6

TOTAL AREA IN LAND AND IN FARMS, U.S., BY AGRICULTURAL CENSUS YEARS 1920, 1945[a]

(1)	(2)	(3)	(4)	(5)
Year	Total Land Area (Acres)	Total Area in Farms		Total Farms (Number)
		(Acres)	(Per cent)	
1945	1,905,361,920	1,141,613,510	59.9	5,859,169
1940	1,905,361,920	1,060,852,374	55.7	6,096,799
1935	1,903,216,640	1,054,515,111	55.4	6,812,350
1930	1,903,216,640	986,771,016	51.8	6,288,648
1925	1,903,216,640	924,319,352	48.6	6,371,640
1920	1,903,215,360	955,883,715	50.2	6,448,343
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

[a] Data from published reports of the Bureau of the Census.

is regarded by the Bureau of the Census as a farm enterprise (that is one on which information is obtained) if it is *on* a "farm"; if such land is not on a farm it is not reported in the "agricultural census." Land under grass is similarly treated—for example "land used under a grazing permit is not to be included" as part of the areal extent of a farm. Even land that contributes directly to the production of food such as "production of maple sirup or sugar with no agricultural operations"

and "picking or gathering of wild nuts, wild fruits, or wild plants (medicinal, ornamental, etc.) except where the land is maintained primarily for their production" is not reported in the agricultural census "unless there are *agricultural* operations". (Italics mine.)

"Farming, or agricultural operations", according to the Bureau, "Consists of the production of crops or plants, vines and trees (excluding forestry operations) or of the keeping, grazing, or feeding of livestock for animal products (including serums), animal increase, or value increase". This seems like a reasonable definition of *agriculture*—except for the exclusion of forestry operations—but by attempting to make this concept synonymous with *farming* the Bureau has made some rather curious decisions, some of which have just been mentioned. The resulting census appears to be one of *farming* rather than of *agriculture*, and should be so labelled. Or, as an alternative, the scope could be broadened to include *non-farming* agriculture[7] as well. By this means the statistical no-man's land, comprising 40% of all United States land, can be brought under adequate statistical description. By bringing these two segments under one comprehensive periodic census it will be possible to eliminate the unknown duplications, omissions and other difficulties to which the present arrangement is subject when the Forest Service, Bureau of Agricultural Economics and other agencies attempt independently to get information on the neglected 40%.

Inadequate unit for presentation of data for small areas. Although it does not publish its agricultural census data on geographic areas smaller than a county, the Bureau of the Census will provide data by Minor Civil Division if and when someone is willing to pay for it. The smallest unit on which the Bureau can provide data is the enumeration district, which is usually the Minor Civil Division (township, beat, etc.) or some subdivision thereof. The Minor Civil Division is of course a unit determined primarily for political purposes. In many states its boundaries are difficult to determine and are short-lived. They do not necessarily describe areas having local significance such as a community.

There is a great need for agricultural data on a small area basis which can be effectively used for land use experimentation, research and planning. There is not only a need for data for areas smaller than a county but also for areas larger than a county but smaller than a state, e.g., the irrigated valleys of the west.

It is suggested that for agriculture a "natural" area be adopted in-

[7] And include all animal and plant production including fish and a number of other categories presently omitted by the Bureau of the Census.

stead of the Minor Civil Division. This area might be based on that which has a common drainage, such as a valley. These will be permanent and will have local significance. They could be the bricks out of which valley authority areas could easily be constructed. They would not interfere with statistics for counties but would greatly facilitate the compilation of useful valley data.

Summary of proposals. A number of inadequacies in the census of agriculture have been discussed, using the 1945 census as the main illustration. Some of these were traced to difficulties of concept such as the conflicts between that which is *agricultural* and that which is *farming* and the operator as a *person* versus the operatorship as an *economic function*. Improvement can be made here by proper labelling and by broadening the scope.

To improve the completeness of enumeration it is suggested that the questionnaire be constructed suitable for use with all households whether they be associated with farms or not. In the open country all households should be canvassed but in cities and towns the canvassing should be confined to a sample (in the agricultural census taken above) in order to avoid heavy costs.

It is suggested that in place of the Minor Civil Division a natural area based on drainage be adopted as the basic unit for providing information on small areas.

THE EDGE MARKING OF STATISTICAL CARDS

A. M. Lester
Montreal

The system of edge-marking of statistical cards described by Dr. Thurstone in the September issue of the Journal has an additional application of considerable importance which he did not mention. This was discovered in the work of analyzing flying accident causes during the recent war. The main data for each accident could be expressed in from 100 to 200 "Yes-No" answers. The total number of accidents to be dealt with was sufficient to make a fully mechanical punch-card system a reasonable possibility, but it was found that reference to a brief written history of each accident was so frequently needed during analysis that hand-sorting of cards which could contain such histories was superior to full mechanization. Cards were utilized with edge punching and the needle was used for most sorting operations, particularly, of course, for replacing the cards in their filing arrangement.

The usual objections to the punching and pinning method were found, however, i.e. errors of punching were difficult to repair satisfactorily and virtually involved making out a new card each time; the edges of the cards eventually tended to tear after much use, and to pin through several thousand cards in search of an element occurring only a very few times was a laborious procedure which might be unreliable owing to cards sticking in the pack.

A system of marking the rims of the cards instead of punching the edges was therefore tried, meaning the actual thin rims of the cards and not the one-eighth of an inch of the front nearest that rim. Fairly thin cards were being used which ran at about 100 to the inch, and it was found not only that a tiny mark on the rim of one of them made with the back of a pen using India ink or red ink was easily visible (when the cards were carefully blocked so that their rims presented an even surface) but also that quite small deviations in the position of the marks could be at once detected. The rims on some types of card were too spongy and caused the ink to run as blotting paper would, but with good quality cards and a fine pen a mark could be made quickly and accurately.

This form of rim marking does not have the advantage of Dr. Thurstone's method, that both sides of the card can be used. On the other hand it can claim the following points in its favor as compared with Dr. Thurstone's scheme, or as compared with the ordinary edge punching scheme:

(1) *ease of reference*—a count can be made while the cards are in their filing drawers if they are tapped into alignment to ensure accuracy;
(2) *ease of extraction*—cards can be picked out without disarranging the other cards. For many types of analysis this is a particularly valuable asset;
(3) *ease of correcting errors*—ink marks on the rims of cards can be bleached or whitened over just as they can on the front of a card but they also can be shaved off with a sharp pair of scissors without seriously impairing the trim of the cards;
(4) marking of the rims of the cards can be used in conjunction with edge punching and with Dr. Thurstone's type of edge-marking.

This type of rim marking might be of value in many statistical analyses although, of course, it has its limitations. It is particularly applicable where, as in accident statistics, a large number of different elements may occur only a few times each in a large universe. In this type of statistical analysis one of the most frequent questions to be answered is how often a certain element occurred in several thousand instances, and what were the circumstances in each case. If the occurrence of the element in question has been marked on the rim of the cards, it is a matter of a few minutes to run through the drawers and pick out the cards in question for further study. For this type of analytical procedure, rim marking would often be quicker than mechanical sorting from punched cards, as well as being enormously cheaper.

CONRAD ALEXANDER VERRIJN STUART (1865–1948)

Another notable statist, honorary member for thirty-five years of the American Statistical Association and outstanding member for fifty years of the International Statistical Institute, has become a part of the history of his major field of work. His friend and compatriot, H. S. Methorst, who succeeded him as Secretary General of the Institute in 1911 and held that post for a quarter of a century, will write for its Bulletin an obituary note based on a wealth of information personal and professional far greater than mine.

It remains for me to outline those aspects of Verrijn Stuart's life which will interest especially the readers of this Journal and accounts of which have been printed in other languages than Dutch. I knew him as a fellow member of the Institute for half a century during which period he missed not one of its twenty biennial sessions (1899–1938); in those sessions he played a leading role. In 1899 he became director of the Central Statistical Bureau of the Netherlands the reorganization of which he described at that time in the Institute's Bulletin and some twenty years later for our Association in Koren's "History of Statistics." In 1902 he printed in the Bulletin a paper on Birth Rates, Still Births and Infant Mortality in certain Dutch Cities and County Districts. Nine years later as Secretary General of the Institute and Chairman of the Committee which organized its Thirteenth Session he welcomed its members and guests to the Netherlands and The Hague.

After the first part of the second Thirty Years War he presented to the Institute as its Rome session (1925) a note on the Representative (Sampling) Method and five years later at the Tokio session a Report on National Capital and Income in the Netherlands. In the following year he submitted to the Financial Committee of the League of Nations a Memorandum on the Gold Question.

In the Revue of the International Statistical Institute which began in 1933 he published a paper (1935) on the Causes of Death and another (1938) on Cancer in the Netherlands. Of more permanent interest probably are his reports on the Mexico City (1933), London (1934) and Athens (1936) sessions of the Institute. The suggestion which he made and repeated in those reports looking towards an improvement in the scientific character of the sessions grew out of an experience almost unparalleled in length and depth, during four years of which he had been Secretary General and then for twenty-seven years an independent friend and critic of his successors, a suggestion

in which I heartily concurred. He thought that the work of the Institute would be improved by a drastic reduction in the number of topics considered at a session, a reduction in the number and importance of section meetings, and a concentration on a few subjects to be probed in the general sessions.

This suggestion was in line with changes made at the Washington session in 1947, changes far more radical than those to which he had looked forward ten years before. We may hope indeed that the forward steps in the progress of international statistics then taken will be found a generation hence to be comparable in importance with those taken between 1875 and 1890 when the Institute arose out of the ashes left by the Franco-Prussian War which had killed its predecessor, the International Statistical Congress. It was for some such an advance that Verrijn Stuart worked and hoped.

WALTER F. WILLCOX

PROCEEDINGS

108TH ANNUAL MEETING

HOTEL STATLER, CLEVELAND, OHIO

MINUTES OF THE ANNUAL BUSINESS MEETING

The American Statistical Association convened for its 108th Annual Business Meeting on the evening of December 28, 1948, at the Hotel Statler in Cleveland, Ohio. A motion was made and passed to approve the minutes of the last Annual Business Meeting held in New York at the Hotel Commodore, December 29, 1947.

Aryness Joy Wickens, Chairman of the 1948 Committee on Fellows reported concerning the fellows elected by the Committee.*

George W. Snedecor announced the appointment of W. Edwards Deming as the new member of the Committee on Fellows to serve for a period of five years from January 1949 through December 1953. Merrill M. Flood gave the reports of the Secretary and Treasurer.*

Walter A. Shewhart, one of the retiring Directors of the Association read the report of the Board of Directors on activities for 1948.* This report had been approved by the Board at its meeting on Monday, December 27, 1948.

Isador Lubin announced that the ASA and the American Economic Association would participate jointly in a memorial service for Wesley C. Mitchell to be held at the Hotel Cleveland, Wednesday, December 29, 1948.

Joseph Berkson reported approval of the activities of the Biometrics Section during 1948.

Isador Lubin read the minutes of the Commission on Statistical Standards and Organization and asked that the recommendations be presented to the membership for such action as they might wish to take.

In the discussion following the motion to accept Mr. Lubin's report, Theodore Brown stated that he thought the report opened the way for, or even by interpretation directed, the Statistical Association to investigate any situation in which it did not approve of the statistical results. Mr. Brown said that he had not the slightest objection to a Commission which would set up standards of good quality in statistical methodology or of ethics in their use, but that in general he believed the work of the Commission should be strictly limited to the setting of such standards.

He indicated that the Commission might well act as an arbitrator in a situation in which a dispute had arisen involving either statistical procedures or their interpretation, but the Association participation, if wise, would be limited to those cases in which the request for a review of a situation originated outside of the Association.

He took very definite exception to the report which seemed to him, from the single reading of Mr. Lubin, to imply that the Association might, on its own initiative, attempt to investigate any statistical procedure in private industry which in the opinion of the committee, or the members representing the Association, or of the Association itself was deemed to be bad statistical practice.

* See the report of the Committee in this issue of the *Journal*

He stated that, "Such police investigations into the work or actions of private business cannot be carried out within the American system of free enterprise and might react so as to bring serious harm to the Association."

Phillip J. Rulon moved that Mr. Lubin's report be accepted without implication that its recommendations be carried out. In further discussion of the motion Harold Hotelling requested clarification of what ASA action would result in "regimentation of private enterprise." Isador Lubin read the following excerpt from the annual membership meeting at Atlantic City in 1947. "The report of the Committee on Standards was approved after some discussion and recommendations that the standing committee to be appointed be expanded to include subcommittees of Association Fellows who would work on the development of standards." Mr. Brown maintained that while he agreed with the original vote of 1947, he did not agree with the proposed recommendations.

Milton Epstein proposed an amendment to the motion which would provide that the words "without implication" be striken from the motion. Theodore H. Brown moved that the entire matter be referred to the Council instead. The Epstein motion was put to the vote and was carried by a substantial majority. Morris M. Copeland moved that the motion be amended to provide that the report be referred to the Council without recommendation. This motion was carried by a standing vote.

President Snedecor expressed his great appreciation of the work of the various committees carried on during 1948, and expressed his particular gratitude to the Committee on Nominations for its excellent work in preparing the slate of officers.

Gertrude M. Cox, Chairman of the Committee on Elections, read the official report of the Committee.*

Simon Kuznets, newly-elected President of the Association, accepted the chair, and as his first official act, gave the floor to George W. Snedecor, retiring president, who delivered his presidential address.**

President Kuznets expressed the gratitude of all members of the Association to George Snedecor for his active leadership during 1948.

Merrill M. Flood announced that the 1949 Annual Meeting would be held in New York, December 27–30, 1949.

Helen M. Walker, reporting for a Resolutions Committee consisting of herself and Morris M. Copeland, presented the following four resolutions:

RESOLVED: That the officers and members of the American Statistical Association express to Gale Ober and the members of his committee their sincere appreciation for the careful planning of all the arrangements for these meetings.

RESOLVED: That the officers and members of the ASA express deep appreciation for the excellent program prepared by members of the Program Committee, namely: Joseph Berkson, Ernest Blanche, John Cover, A. Ford Hinrichs, Simon Kuznets, Rensis Likert, Abraham Wald, Allen Wallis, Harry Wellman, Aryness Joy Wickens, and Merrill M. Flood, *Chairman.*

Whereas the Placement Committee of the New York Chapter has, during the past year, carried on a remarkably effective service to its members, RESOLVED: That the Board and Council give serious thought to ways in which such placement service can be more adequately financed and can be extended on a nationwide scale.

---

* See Report of the Committee, in this issue of the *Journal.*

** March, 1949, *Journal of the ASA.*

Whereas those members of the ASA who have major interest in the social sciences constitute a large proportion of the present membership of the Association, and whereas many of these members have expressed dissatisfaction with the benefits received from their membership, RESOLVED: That the Board and Council give serious thought to ways and means of providing a balanced program of services to all its members.

These four resolutions were voted on by the membership and approved.

The following two resolutions were transmitted by Miss Walker without any recommendation from her Committee.

Resolution on Marriage and Divorce Statistics: Whereas, the American Statistical Association recognizes the need for more adequate marriage and divorce statistics, to provide data needed in statistical research in many areas such as demography, health, business, welfare and others; and

Whereas the development of national vital statistics of marriages and divorces depends upon cooperative state-federal relationships along the lines which have proven effective for the vital statistics of births and deaths:

RESOLVED, that the American Statistical Association recognizes the need for state centralization of marriage and divorce records and statistics and for their integration with other vital records and vital statistics; and urges all states not yet operating such an integrated system to initiate such a program at the earliest possible time; and

RESOLVED, that the American Statistical Association calls upon the Federal Security Agency, through its National Office of Vital Statistics in the Public Health Service, to encourage the development of systems of marriage and divorce records and statistics in every state leading to maximum comparability and prompt availability of data; and to undertake the development of detailed national marriage and divorce statistics, adequate to meet the pressing needs for such vital statistics.

Resolution on Birth Statistics: Whereas, birth statistics are essential to developing local, state, and national estimates of population changes, determining trends in family size, and are required for planning and evaluating health, welfare, and related programs, and

Whereas, the usefulness of birth statistics would be increased by improved registration completeness and by a current knowledge of the completeness in which births are registered in local areas and by various characteristics of the population, and

Whereas, such a knowledge would also provide the basis for promotion of complete registration,

RESOLVED, that the American Statistical Association strongly urges that a uniform, nation-wide test of registration completeness be carried out in conjunction with the 1950 decennial census of population in order to provide the required measures of registration completeness on local, state, and national levels and that every effort be made to obtain more complete registration.

The membership voted to refer these two resolutions to the Council for consideration.

Aryness Joy Wickens rose to present the following resolution.

RESOLVED, that a special Sub-Committee be set up by the Committee on Committees to consider and formulate specific standards for the selection of Fellows, in carrying out the constitutional provision that fellows shall be "statisticians of established reputation"; that the recommendations of this special

Sub-Committee be forwarded to the Board and Council for its approval; whereupon the approved standards shall constitute the criteria for the selection of fellows.

Theodore H. Brown, Robert W. Burgess, and Stuart A. Rice spoke for the resolution. W. Edwards Deming took the position that the Committee on Fellows should review its own standards. The resolution was put to the vote and was passed in the form in which Mrs. Wickens presented it.

Eugene Pike introduced a resolution which would provide that the Council, where it is in the interest of the membership, state those matters of policy before it and ask the membership to communicate with the Council concerning it. Milton Epstein, in discussing the Pike resolution, asked that one-half day of each Annual Meeting be set aside for business meetings so that the membership might have an opportunity to prepare resolutions and discuss all matters of policy before the Association. The resolution was voted as presented. Helen M. Walker reminded the assembled members at this point that annual business meetings are not mandatory under the new Constitution and that any such provision would have to be made on the initiative of the officers of the Association.

The meeting was adjourned.

Report of the Board of Directors

The year 1948 has been one of transition in the affairs of the Association. It has been marked by the orderly movement toward the organization structure envisaged in the new Constitution, by the reorganization of the national office under a new Secretary-Treasurer, and by an unavoidable increase in membership dues.

The new Constitution was adopted by mail ballot of the membership in February 1948 and became effective on January 1, 1949. Nominations and elections of officers for 1949 were in accordance with the new Constitution. For the first time in the history of the Association the membership voted by mail ballot for national officers. The representatives of the newly formed Districts were also elected to the Council. The Council, now the policy forming body of the Association, held its first meeting on December 29, 1948.

The 108th Annual Meeting of the Association was held in Cleveland between December 27 and 29. Joint sessions were held with the American Economic Association, the American Marketing Association, the Econometric Society, the American Farm Economic Association, the Population Association of America, the American Public Health Association, the Institute of Mathematical Statistics, the Biometric Society, and the Ohio Section of the American Society for Quality Control. Over 600 people attended the sessions.

It is noteworthy improvement that the planning of the 1948 Annual Meeting program was completed before October. Moreover, the 1949 Annual Meeting program has already been outlined, approved by the 1948 Program Committee, and referred to the 1949 Program Committee for consideration and further planning.

Definite arrangements have been concluded for a four-day meeting in New York City during the Christmas holidays of 1949. The fortunate circumstance that both the American Association for the Advancement of Science and the Allied Social Science Associations will be meeting in New York at that time made it possible for the Association to plan combined sessions that will cover an exceptionally broad subject field. Joint sessions have already been scheduled

with the American Economic Association, American Farm Economic Association, American Marketing Association, American Psychological Association, American Society for Testing Materials, Biometric Society, Econometric Society, Institute of Mathematical Statistics, Population Association of America, and the Housing Research Committee of the Social Science Research Council. There are tentative plans for other joint sessions with the American Society of Mechanical Engineers, American Astronomical Society, and the Psychometric Society.

The situation with respect to Association membershp is not entirely heartening. It will be recalled that the Board of Directors reluctantly recommended a substantial increase in membership dues from $5.00 to $8.00 per annum, the increase being financially unavoidable in order to meet continually increasing costs of printing and other services in an inflationary period. This decision was ratified by mail ballot of the membership in February 1948. The withdrawal of a portion of the membership in the face of this dues increase, either because of a marginal interest in statistics or because their real incomes were adversely affected by the inflation, had, of course, been anticipated and discounted in advance. The actual result for 1948 is that the size of the Association has remained about constant. About 800 members had dropped out and an about equal number of new members have come in. While this withdrawal of 800 members is regrettable, it is surely not a recurrent item. During the year, new membership applications totaled approximately 725 as against 1,100 in 1947. This slackening of the rate of increase can probably be attributed to the deterrent effect of the higher dues in this and other Associations, and to the fact that vigorous canvassing in 1946 and 1947 of the most likely sources of new members has forced recourse to more marginal groups from the standpoint of their concern with statistics.

The main organizational change in the national office has been the appointment of a new Secretary-Treasurer after the resignation of Dr. Lester S. Kellogg from that position at the 1947 Annual Meeting. Sylvia C. Weyl took over as Acting Secretary-Treasurer on an ad-interim basis until May. Durmg the spring of 1948 an ad hoc nominating committee, under Samuel S. Wilks, energetically searched the field for a man combining the many managerial and scientific attainments necessary for the executive leadership of the Association. After considering a number of candidates Merrill M. Flood was chosen, and his nomination as Secretary-Treasurer was unanimously approved by the Board of Directors. In addition to his duties as Secretary-Treasurer, Dr. Flood served as Chairman of the 1948 Program Committee which organized the 1948 Annual Meeting and planned the sessions to be held in 1949. The Councıl of the Association at its December 29 meeting re-elected Dr. Flood for a 3-year term.

Two new chapters were added to the Association in 1948. A Saint Louis Chapter is now functioning regularly, and the old and established Sacramento Statistical Society has become a chapter of the Association.

A temporary Committee on Publications was appointed by President Snedecor, with William G. Cochran as Chairman to outline the policies to be followed by the various Association publications, to suggest better means of cooperation among them, and to make other such recommendations within this general area as it deemed proper. This Committee has now tendered its report to the Board of Directors. Under the new Constitution, the standing Committee on Publications will continue this function.

The Commission on Statistical Standards and Organization, under the chairmanship of Isador Lubin, was organized and held its first meeting in November.

Three of its members participated in the survey by the Social Science Research Council of public opinion polling methods—with special reference to the inaccurate forecasts of the 1948 election returns. Several problems of major importance have been placed before the Commission. It will surely be in a position of influence and develop rapidly during 1949 as a positive force for the advancement of statistics.

With the 1950 Census of the United States already approaching, the Census Advisory Committee has been actively engaged in aiding the Director of the Census in reviewing and appraising schedules. It has screened many requests from the public to the Census Bureau for inclusion of additional questionnaire items. This Committee has an added responsibility now that sampling methods are under fire and yet of such critical importance in the work of the Bureau of the Census.

Within the field of Association publications, the Biometrics Bulletin has been increased to four 64 page issues, and both format and contents have been greatly improved. The name of the publication has been changed to Biometrics. The American Statistician, launched in the summer of 1947, has been very well received by the membership and has already established itself as a valuable periodical of the Association.

The Association cooperated informally in the work of the National Bureau of Economic Research under Frederick C. Mills, for the Hoover Commission on the Organization of the Executive Branch of the Government. At the request of Dr. Mills, President Snedecor arranged for an independent study of the statistical activities of the Federal Government and preparation of a report for the member group. John W. Tukey canvassed a representative group of members of the Association, many of whom are familiar in detail with the statistical practices and problems of the Federal Government, and obtained their evaluation of the caliber of federal statistical work and their recommendations on its consolidation, extension and improvement. His excellent report, which summarizes the judgments obtained was forwarded to Dr. Mills.

The International Statistical Institute advised the ASA of the adoption of its new statutes, which provide for the affiliation of national and international statistical societies. The ISI has been reorganized to work as an active professional organization dedicated to the furtherance of statistical science throughout the world. In view of the similarity of purpose of the Institute and the Association, Stuart A. Rice, president of the ISI, expressed the hope that the Association would request affiliation. The proposal is being considered by the Council.

The Association has an increasing interest in the applications of statistics to management problems, many of its members are active in such fields as business economics, marketing, and quality control. The National Management Council, a federation of national organizations concerned with management problems has invited the Association to become a member. Among its present members are the Society for the Advancement of Management and the American Society of Mechanical Engineers. The invitation will be acted upon by the Council.

The Eastern North American Region of the Biometric Society requested affiliation with the ASA during the past year. Decision on this application will be made by the Council. The Biometric Society, an international organization of biometrics workers, decided in 1948 to designate Biometrics as its official organ. Block subscriptions to Biometrics were made available to members of the society at a reduced rate as a means of assistance for the new Society.

Plans for several new activities of the Association have been partially de-

veloped during the year. Among these are establishment of a speaker's bureau to serve the chapters, formation of a research group to study the program and operations of the Association, and intiation of regional District meetings. Some or all of these new projects will be activated during 1949. These new projects are an integral part of your Board's program during 1949.

Regrettably, the need for economy during 1949 made it necessary to postpone indefinitely the project for launching the periodical "Statistical Reviews."

The past year has been one of unavoidable financial stress and structural reorganization. The Board considers that the major problems involved in this postwar reorganization have been faced and are now on the road to solution. The task ahead of the Association is to consolidate organizationally and financially and at the same time to continue to expand its membership and carefully selected activities, in the face of an inflationary trend that is prejudicial both to professional societies and to the memberships they serve.

GEORGE W. SNEDECOR, *President*
ISADOR LUBIN
LOWELL J. REED
WALTER A. SHEWHART
FREDERICK F. STEPHAN
SAMUEL STOUFFER
WILLARD L. THORP
MERRILL M. FLOOD, *Secretary-Treasurer*

The Secretary's Report on Membership

There were 4,720 members remaining at the beginning of 1948 after 189 were dropped for non-payment of dues. Although 725 new members joined during 1948, and 27 were reinstated, there was a net decrease of 70 bringing the total on December 31, 1948 to 4,650.

The 1948 membership is composed of the following groups:

Honorary members	13
Fellows	138
Student members	290
Regular members	4,209
Total Membership	4,650
Corporate Members	6

Over 200 members resigned during the year and some 600 were dropped for non-payment of dues. The primary cause for this heavy loss in membership is undoubtedly the increase in dues from $5.00 to $8.00. The unfortunate delay in settling the matter of dues increase, and the consequent lateness in getting dues notices to the members, also had an adverse effect on membership.

The members of the Biometrics Section, at the end of December 1948, number 960, of whom 170 are associate members and 790 are also regular members of the Association.

The deaths of the following members were recorded during the year: Wesley C. Mitchell, *Fellow;* Edward G. Benson, Bertram Butler, Ruth Dawson, Valentino Dore, T. Bertrand Graham, Edward W. Higgins, Arthur Hurd, Walter E. Magney, Douglas W. Oberdorfer, Hallie K. Price, O. A. Pope, H. M. Tompkins, John H. Watkins, *Regular Members.*

MERRILL M. FLOOD, *Secretary*

Report of the Nominating Committee

The Nominating Committee of the American Statistical Association announces the election of the following officers for the year 1949.

President	Simon Kuznets
President Elect	S. S. Wilks
Vice President	
3 years	Dorothy S. Brady
2 years	H. A. Freeman
1 year	Lester S. Kellogg
Directors (3 year term)	
	C. H. Goulden
	L. L. Thurstone
District Representatives	
Western District	Maurice I. Gershenson
	Henry B. Moore
North Central	Howard L. Jones
Southeastern	Samuel Weiss
	Morris H. Hansen

They were selected as persons

(1) who are distinguished in their contributions in applying methods to problems in various fields
(2) who are acquainted with and contributing to statistical theory
(3) who represent the various fields of membership interests and the different geographical regions of the country.

Respectfully submitted,
Nominating Committee:
Milton Friedman
Charles F. Sarle
Mortimer Spiegelman
Holbrook Working
Gertrude M. Cox, *Chairman*

Report of the Committee on Fellows

The American Statistical Association has announced the election of four of its members as Fellows—three American statisticians and one Canadian statistician. (There are now 145 Fellows of the American Statistical Association from a total membership of about 4,700.)

The newly elected Fellows are:

Eugene L. Grant, Professor of Civil Engineering at Stanford University, who has developed special courses for teaching statistics for engineers, and who has been a leader in the practical application of statistics to engineering.

Tjalling J. Koopmans, of the Cowles Commission for Research in Economics at the University of Chicago, who has done outstanding work in the application of statistics to economics.

W. Allen Wallis, Professor of Statistics and Business Economics at the School of Business at the University of Chicago, who was Administrative Director of the Statistical Research Group of the Office of Scientific Research and Development which developed and supplied answers to many statistical problems of military importance submitted by the armed services.

J. W. Hopkins, Canadian biometrician, of the Division of Biology and Agriculture of the National Research Laboratory of Canada and Coordinator of the Special Committee on Applied Mathematical Statistics, of the National Research Council of Canada. His principal work has been done in applying statistical methods to experiments with grains and livestock. During the war he served as scientific advisor to the Air Vice Marshal of Canadian anti-submarine warfare.

MINUTES OF THE MEETING OF THE COMMISSION ON STATISTICAL STANDARDS AND ORGANIZATION

Present: S. Wilks
W. Shewhart
F. Croxton
I. Lubin

Absent: L. Reed

CHARTER RECOMMENDATIONS

The Commission recommends that the Committee on Committees, in preparing the charter for the Commission on Standards, should define its functions and competence in the terms stated in the report approved by the members of the American Statistical Association at its annual business meeting on January 25, 1947, as follows:

> The Committee should have rotating membership based on a three year term; one-third of the initial membership to be appointed for one year, one-third for two years, and one-third for three years. Members should be eligible to reappointment. Election should be made by the Board of Directors after consultation with the Committee.
>
> The functions of the Commission should be:
>
> A— to provide a tribunal to render opinions and recommendations on controversial issues relating to statistical procedure and presentation of statistical material.
>
> B—develop a list of minimum standards for published statistical materials
>
> C—upon request from governmental bodies, review actual or proposed undertakings and make recommendations relative to standards.
>
> The Committee might eventually develop a code of ethical practices in statistical work.

It is recommended that these terms be amplified to the extent of adding at the end of the last sentence above the words "with a view to enhancing the status and acceptance of statisticians by the general public."

Rules of Procedure. The Commission decided to make no recommendations relative to rules of procedure. It was the consensus that these rules should be developed as experience dictates.

Membership. The Commission recommends that Joseph Davis of Stanford University and Samuel Stouffer of Harvard University be added to its membership.

Budget and Staff. It is impossible at this time to make any definite estimate as to the budgetary requirements of the Commission. Its financial requirements will depend upon the amount of work that will be undertaken during the coming year, the nature of the projects it undertakes and the extent to which Governmental agencies will finance studies and surveys made upon Governmental request.

The Commission, however, feels that it is essential that there be a part-time secretary available to it. Eventually it will be necessary to have a full-time secretary. In order to finance a part-time secretary and the expenses of Commission and Sub-Commission members, it is estimated that for the year 1949 approximately $15,000 will be required if the Committee is to undertake projects that are listed below.

Projects

The Commission gave consideration to possible projects that it might undertake in the immediate future.

1. *Joint Committee on the Economic Report.* The Commission had before it a request submitted by the Staff Director of the Joint Committee on the Economic Report of the House and the Senate relative to its reports on "Statistical Gaps" and "Economic Indicators." The Commission felt that this request should be complied with and that it should undertake to study and assess both of these Joint Committee reports. It is proposed to set up a Sub-Commission at an early date to undertake this project.

2. *Consumer Price Index.* The Commission feels that a request from the Commissioner of Labor Statistics to advise the Bureau of Labor Statistics concerning technical questions involved in constructing the Consumer's Price Index be complied with. It was the feeling of the Commission, however, that the success of projects of this sort will be greatly enhanced if its services were requested before any definite action has been taken by the agency involved. In other words, the Commission would prefer that it be called in at the time that technical problems are in process of being studied rather than to pass judgment at a later date upon already determined procedures.

3. *Standards for Publication of Federal Statistical Data.* The Commission agreed that it should comply with the request of the Bureau of the Budget for a review of the Bureau's report on "Standards for the Publication of Statistical Data." It felt, however, that it was not within its competence to comply with the Bureau's request that it suggest principles that should guide the Federal Government in its release or suppression of statistical information during an emergency.

4. *The Adequacy of Present Types of Federal Statistics.* At the April 1948 meeting of the Board of Directors a resolution by Mr. Solomon Barkin to the effect that the American Statistical Association should investigate its responsibilities in the field of industrial statistics was referred to this Commission. At a later date the Labor Advisory Committee of the Bureau of the Budget adopted a resolution submitted by Mr. Barkin requesting that the ASA be called upon to establish a committee to appraise the adequacy of present types of statistical data and the direction in which statistical information should be developed.

The Commission considered these resolutions and is of the opinion that no action should be taken upon them until the report of the Hoover Commission is made public. Considerable time and energy has been devoted by the Staff of the Hoover Commission to the question of the collection and development of federal statistics. Indeed, a paper on the work of the Hoover Commission in the field of statistics will be delivered at the forthcoming annual meeting of the American Statistical Association by Dr. Mills. The Commission recommends that any possible action be deferred until the Hoover Report can be studied and analyzed.

5. *The Kinsey Report.* It has been suggested that the Commission review the Kinsey report with a view to passing judgment on the methodological statistical

techniques used by the author. It was the unanimous opinion of the Commission that we do nothing in this field unless specifically requested by a representative body of American citizens.

6. *Polling Techniques.* The widespread criticism of the election polling results has resulted in requests that the Commission initiate studies of the techniques used by the various polling organizations. In view of the fact that the Social Science Research Council has appointed a special committee to study the election polls and in view of the further fact that two members of this Commission are on the SSRC Committee, it was felt that no action should be taken in this field at this time. The Commission was aware of the fact that further work in the field of polling would have to be done after the SSRC Committee completes its report. However, provision has already been made for such work by the National Research Council and the Social Science Research Council. Arrangements have been made for studying special problems such as sampling, interviewing, and panels. It may be advisable after these studies have been completed for the Commission to look further into the question of polling techniques. At the moment, however, it feels that anything it might undertake would be more or less in the nature of duplicating work already under way.

Independent Investigations. The Commission feels that in those instances where it is not specifically requested by a Government agency, by a foundation or similar body which financed or sponsored a study, or by a representative body of citizens to undertake investigations of projects or of statistical standards it should act on its own initiative only in those instances where the public interest, domestic or international, is concerned. As problems arise that are affected with the public interest, the Commission will take under consideration the advisibility of investigating them with a view to pointing out to the public such weaknesses and inadequacies of statistical techniques and presentation as it may find. Its primary purpose shall be to further higher standards in statistical fields and thereby enhance the status of statisticians in the mind of the public.

Report of the Treasurer

The 1948 budget as originally approved by the Board of Directors planned for an income of $61,900 and expenses of $58,925. Actual income was $51,320 and expenses were $64,605. In September, the Board voted a revised budget providing for the increase in expenses and decrease in income.

Income was less than expected primarily because some 600 more members than had been estimated left the Association during the year. Also there were some 275 fewer new members than estimated, and receipts from subscriptions, sales, and advertising were some $4,000 less than estimated. Income for 1948 exceeded that for 1947 by $14,315, nevertheless, because of the additional revenue derived from higher dues and subscription rates.

Expenses were somewhat higher than anticipated because of increases in printing costs. Printing costs will be still higher in 1949.

The national office plans to continue its drives for new members and to make a new drive for additional advertising and subscriptions. Through this expansion the Association will widen its sphere of influence and increase its services, and at the same time improve its financial status.

MERRILL M. FLOOD, *Treasurer*

Report of the Auditors

To the Board of Directors of
American Statistical Association

We have examined the attached financial statements of American Statistical Association relating to the year ended December 31, 1948. Our examination was made in accordance with generally accepted auditing standards, and accordingly included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

The recorded cash receipts for the year were traced to the deposits shown on the bank statements and the amounts for dues and subscriptions were tested with the membership and subscription records. The paid checks were inspected and related vouchers tested in support of cash disbursements for the year. The bank balances were reconciled with amounts reported direct to us by the depositaries and the cash on hand and the securities owned at December 31,1948 were verified by inspection. We did not check the membership and subscription records in detail or make any independent verification of the inventory of old Journals, the office records of which are based, in part, on data assembled in prior years, no recent physical inventory having been taken.

The life membership reserve at December 31, 1948 reflects the amount needed to support a life annuity for each life member in the same annual amount as that which could have been purchased by the original lump sum payment, based on a 2½% interest rate, the 1937 Standard Annuity Table and the age of the life member when the lump sum payment was made, in accordance with a resolution of the Board of Directors adopted pursuant to a mail ballot in January 1949. Previously the reserve had been calculated on the basis of the combined annuity table of mortality with assumed interest at 4% per annum and an assumed annuity of $5 per member. The amount treated as income from life memberships in 1948 represents the excess of the reserve at the beginning of the year over the required reserve, on the new basis, at the end of the year.

The accounts for the year 1948 include for the first time a provision for employees' accrued annual leave. The reserve of $1,209.54 provided at December 31, 1948 includes $506.50 applicable to the prior year. We understand that this provision was made at the direction of the Secretary.

In our opinion, the accompanying statements present fairly the position of American Statistical Association at December 31, 1948 and the results of its operations for the year, in conformity with generally accepted accounting principles applied on a basis consistent, except as mentioned in the two preceding paragraphs, with that of the preceding year.

PRICE, WATERHOUSE & CO

Washington, D. C.
April 27, 1949

American Statistical Association Balance Sheet

	December 31, 1948	*1947*
Cash in bank and on hand	$ 7,513.00	$ 1,853.91
Accounts receivable	1,204.97	1,356.22
Investments:		
United States Savings Bonds, Series D, at redemption value	6,406.00	6,138.00
Stocks, at cost (at market quotations $8,219 and $6,150, respectively)	5,793.50	5,793.50
Inventory of old Journals, at approximate cost	1,907.51	2,024.63
Furniture and equipment, at cost less depreciation	2,892.19	2,482.76
Deferred charges	277.80	
	$25,994.97	$19,649.02
Liabilities		
Accounts payable	$ 4,631.43	$ 6,515.76
Note payable	5,000.00	
Accrued interest	37.50	
Accrued annual leave	1,209.54	
Deferred income (collections applicable to subsequent year):		
Dues	15,964.53	1,054.50
Subscriptions	3,797.59	2,704.76
Other		115.35
	$30,640.59	$10,390.37
Net worth:		
Life membership reserve	$ 3,131.40	$3,750.20
Surplus, per statement	(7,777.02)	5,508.45
	$ (4,645.62)	$ 9,258.65
	$25,994.97	$19,649.02

Surplus Statement	*Year ending December 31, 1948*	*1947*
Balance at beginning of year	$ 5,508.45	$ 4,248.79
Add—Transfer of balance in Centenary Sustaining Fund		6,379.51
	$ 5,508.45	$10,628.30
Deduct—Excess of expenses over income for the year, per income statement	13,285.47	5,119.85
Balance at end of year	$ (7,777.02)	$ 5,508.45

AMERICAN STATISTICAL ASSOCIATION
INCOME STATEMENT

	Year ending December 31, 1948	1947
Income:		
Dues—current year	$ 32,495.90	$21,298.58
Dues—prior years	54.65	215.00
Life membership income	418.80	397.33
Subscriptions	7,386.28	6,206.95
Advertising	1,728.77	2,010.08
Reprints	892.09	509.00
Journal sales	2,702.19	2,827.80
Biometrics Section income	4,212.70	2,466.86
Miscellaneous	724.15	684.70
Dividends and interest	704.50	398.57
	$ 51,320.03	$37,014.87
Expenses:		
Journal—printing, mailing and reprints	$ 12,160.70	$10,105.81
Salaries and wages, including in 1948 $506.50 accrued annual leave expense applicable to prior year	30,582.81	17,037.26
American Statistician Bulletin	5,875.55	4,259.88
Biometrics Section expenses	4,316.29	2,366.46
Rent	2,880.00	1,365.00
Office supplies, printing and mimeographing	1,599.27	1,764.42
Postage	1,748.74	1,511.11
Telephone and telegraph	754.81	512.03
Travel expense—officers	649.20	603.91
Annual meeting expense	613.31	647.89
Depreciation of furniture and equipment	442.86	324.99
Promotion expense	743.28	
Loss on disposal of equipment		151.09
Storage of old Journals	71.95	72.00
Cost of old Journals sold	573.10	382.27
Miscellaneous	1,593.63	1,030.60
	$ 64,605.50	$42,134.72
Balance, loss, carried to surplus	$(13,285.47)	$(5,119.85)

BOOK REVIEWS

Edited by

OSCAR KRISEN BUROS

Rutgers University

Actuarial Statistics: Vol. II, Construction of Mortality and Other Tables. *J. L. Anderson* (Scottish Widows' Fund and Life Assurance Society, 9 St. Andrew Square, Edinburgh, Scotland) and *J. B. Dow* (Standard Life Assurance Co., 3 George St., Edinburgh, Scotland). Published for the Institute of Actuaries and the Faculty of Actuaries. London N.W. 1: Cambridge University Press (Bentley House, 200 Euston Road), 1948. Pp. xvi, 281. 21*s*.

REVIEW by T. N. E. GREVILLE
Chief, Actuarial Analysis Branch, Public Health Service
Federal Security Agency, Washington 25, D. C.

THIS book has been written primarily to assist British actuarial students in their preparation for the examinations of the Institute of Actuaries and the Faculty of Actuaries. As might be expected, the main emphasis is on the construction of mortality tables from the records of insured lives and life annuitants. However, the two chapters on the construction of national life tables at least acquaint the reader with the more important problems that usually arise in that connection; and the final chapter on sickness rates is also very informative to the uninitiated.

The principal topic, the construction of mortality tables from the records of insured lives, is rather fully discussed, and there is no doubt that this is the most complete treatment of the subject thus far published. The very detailed explanation of exposed-to-risk formulas will be especially helpful. Throughout the book the exposition is lucid and accurate. Clarity is assisted by numerous illustrative examples, which add greatly to the usefulness of this text. The absence of problems to be worked out by the student is a shortcoming which it has in common with other British actuarial textbooks.

It is natural that the principal emphasis should be placed on developments in Great Britain. However, after making all due allowance, it is still the impression of the reviewer that the book suffers somewhat from a certain insularity of approach. Certain useful contributions have been made by actuaries in the United States and Canada which, it is suggested, could most appropriately have been mentioned in a book purporting to deal comprehensively with a branch of scientific methodology. One is surprised, for example, to find no mention of abridged processes of life table construction, a subject in which there have been important developments on this side of the Atlantic, but in which British demographers have expressed considerable interest. Again, on page 255, where it is pointed out that occupational mor-

tality statistics based on national population data "are not reliable guides for a life office to use for assessing extra premiums for occupation," some mention might have been made of the extensive studies of occupational mortality among assured lives in the United States and Canada made by the Actuarial Society of America and the Association of Life Insurance Medical Directors.

This is, of course, a minor criticism; and it is the reviewer's opinion that the authors have, on the whole, handled their material in expert fashion and have produced a text which will be most useful not only to the students for whom it is intended but also to any others having an interest in the subject matter.

Experimental Designs in Sociological Research. *F. Stuart Chapin* (Professor of Sociology, Chairman of the Department, and Director of the School of Social Work, University of Minnesota, Minneapolis, Minn.). New York 16: Harper & Brothers (49 East 33rd St.), 1947. Pp. xi, 206. $3.00.

REVIEW BY MARGARET JARMAN HAGOOD
Statistician, Division of Farm Population and Rural Life
Bureau of Agricultural Economics, Washington 25, D. C.

IN THIS brief volume, Dr. Chapin has summarized the designs and findings of nine sociological "experiments" and, in addition, has presented a considerable amount of general exposition of experimentation in sociology and of sociometric scales. As Chapin points out, the work is complementary to, rather than in competition with, Ernest Greenwood's *Experimental Sociology*, which appeared two years earlier. Greenwood treated the subject with primary attention to the conceptual formulation and logical principles involved, with a critical examination of the literature. Chapin, on the other hand, had as his purpose to "illustrate the method of experimental design by reproducing concrete studies" and to provide "a source book of examples of specific application analyzed in some detail" (p. ix).

Dr. Chapin has fulfilled this purpose excellently—in fact, no other person in the United States is in position to fulfill it equally well. To Chapin goes the credit for leadership among sociologists for trying to adapt the experimental approach to sociological research problems and for decades of work and encouragement of work of others in this field. He and his students have made outstanding advances in quantification of hitherto unmeasured social phenomena and in continuously carrying out pioneering work to establish sociology as a science.

Without reflection on the author, one can differ with him, as to when the term "experiment" or "experimental" is to be validly used. I am inclined to be much more restricted than is Chapin when it comes to classifying a social research project as an "experiment." The social "experiments" reported in his book do not measure up to the criteria that I hold for "experiments."

But I do agree with Chapin that in our field, the research worker is unrealistic who writes as though one could control the "treatments," and provide for random selection of the units which are to receive "treatments," as in the case of biology and in some areas of psychology.

R. A. Fisher's methods of analysis of variance and covariance have not yet been fully explored in comparison with Chapin's matching technique which sacrifices much data. Dr. Chapin is aware of this and it is hoped that he will focus research on the problem in the future.

Elements of Mathematical Statistics. *C. V. L. Charlier.* Including **Table of Poisson's Function** by *L. v. Bortkiewicz.* Edited and translated by *J. A. Greenwood.* Brooklyn 25, N. Y.: J. A. Greenwood (25 Winthrop St.), 1947. *Two reviews follow:*

Review by Burton H. Camp
Emeritus Professor of Mathematics, Wesleyan University
Middletown, Connecticut

This well-known book, written in German and published in Hamburg in 1920, is now available in an English translation, and it is this translation which is the subject of this review. The translator has also made occasional emendations of the text and an improvement in the tables. He himself has computed new tables for ϕ_2, ϕ_3 and ϕ_5, these functions being derivatives of the function defining the normal curve; he has made some corrections in the original tables; and has added a new table, No. 44. The meanings of Tables 43 and 44 are not indicated in the appendix where the tables occur. It can be learned from the preface that Table 43 is the table of Bortkiewicz, that is, a table of the Poisson function, the reader being left to infer the meanings that must be attached to the numbers at the top and side of the table. There is no easy way of finding out what Table 44 does mean, but the thorough reader will find it explained in a footnote on page 60.

So far as the reviewer can judge the translation is sufficiently accurate. The English is often awkward but not without a quaint charm so that perhaps by its very defects the phraseology makes one the more aware of the originality of the distinguished author who is thus being presented to us. In fact one would read Charlier's book now more to become acquainted with Charlier's way of looking at things rather than to study the material which the book contains. This material has been worked over many times since 1920 and has been presented in various forms in several books; but there is much value nevertheless in going to the original source, and this translation makes it possible for the English reader to do this. The general scope of Charlier's book may be indicated as follows. First of all there is a discussion of homograde or alternative statistics. These are statistics showing the number of times an event occurs, in a given number of trials, when it may either occur or not occur. This sort of frequency distribution may be

approximated by successive terms of the point binomial, in certain cases, and in other cases by the modifications of the point binomial introduced by Poisson and by Lexis. Heterograde statistics comprise frequency distributions that are now sometimes called continuous. These may be approximated sometimes by the normal probability curve, sometimes by the more general A-Type or B-Type curve. The subject of simple correlation is considered at some length, and the discussion includes what Pearson calls tetrachoric correlation.

Review by Alexander M. Mood
Statistician, Project RAND, Douglas Aircraft Co., Inc.
1500 Fourth Street, Santa Monica, California

It is always interesting to read an older book and compare the state of statistics then with now. The present little volume was written in 1910 (the translation was made from a 1920 German translation) just two years after Student's famous paper was written and before the full import of that paper was recognized. Fisher's paper on the foundations of statistics was not to appear for ten more years.

The book is divided into two parts of about fifty pages each; the first part discusses what would now be called discrete distributions, and the second, continuous distributions. The discussion is largely in terms of the first four moments of the distributions and of course in the second part there is considerable attention to the Gram-Charlier system of frequency functions. Actually there is much in the book that can be found in the best sellers in statistics today, but unfortunately that fact cannot be interpreted as a compliment to Charlier's sagacity—it is merely an indictment of our best sellers. I refer to such things as the detailed computational instructions in terms of class frequencies all outmoded by modern computing machines, the hoary errors about skewness and kurtosis still fed our students today, the confusion between population parameters and their estimates, estimation by the method of moments only, the great faith in the standard error whether the distribution is normal or not. But Charlier, at least, cannot be much blamed for these shortcomings because they were either not relevant or had not been pointed out at the time.

The real trouble with the book is simply that it was written before the Fisherian revolution in statistics. It was written in that period when statisticians were trying to develop and formalize statistical inference in terms of relatively simple classes of frequency functions—the Pearson curves and the Gram-Charlier series. Looking back, it is easy enough to see that such a course was actually a blind alley, that a fitted curve contains no more information than the data to which it is fitted. But at the time that approach must have looked promising and Charlier did significant work in the field.

Today the book is of little interest except possibly to students of the history of statistics, and the reviewer is at a loss to rationalize the presentation of this translation.

Elements of Nomography. *Raymond D. Douglass* (Professor of Mathematics) and *Douglas P. Adams* (Assistant Professor of Graphics). (Massachusetts Institute of Technology, Cambridge, Mass.) New York 18: McGraw-Hill Book Co., Inc. (330 West 42nd St.), 1947. Pp. ix, 209. $3.50. [London W.C. 2: McGraw-Hill Publishing Co., Ltd. (Aldwych House, Aldwych), 1948. 21s.]

REVIEW BY JOSEPH ZUBIN
Associate Research Psychologist, New York State Psychiatric Institute
722 West 168th St., New York, N. Y.

THIS is an elementary text of nomography for students in engineering which may serve very well as an introductory text for any student of statistics who wishes to know how nomographs may be constructed. From the point of view of the statistician it is unfortunate that no direct applications of nomography to statistical problems are made. Perhaps that is a task for a future nomographer who has statistics well in hand. This simple elementary text ought to prove useful for those students who wish to become acquainted with the use of nomographs in their field.

The more routine tasks of computation such as comparison of percentages and comparison of correlation coefficients with their standard errors are more readily accomplished by means of nomographs than by calculating machines and tables. There are several source books in which such nomographs are provided but the manner of their construction and limitations on their use is not always understood by the average statistician. Furthermore, no new statistical aids of this type can be invented unless statisticians become aware of the problems involved in nomograph construction and of the areas in which they may prove useful. The simplicity of the present text and its empirical approach ought to win adherents for the use of nomographs on a wider scale.

A Guide to Public Opinion Polls, Second Edition. *George Gallup* (Director, American Institute of Public Opinion, Princeton, N. J.). Princeton, N. J.: Princeton University Press, 1948. Pp. xxiv, 117. $2.50. [London E.C. 4: Oxford University Press (Amen House, Warwick Square). 14s.]

REVIEW BY ROBERT COBB MYERS
Educational Testing Service, Princeton, N. J.

THIS is a revised edition of the question-and-answer handbook which was first prepared in 1944. The avowed purpose of the book is "to answer, in non-technical language, the questions that people most frequently ask about public opinion polling." If *non-technical* can be considered as synonymous with *naive, unsophisticated, incomplete, repetitious,* and *contradictory*, we can then concede that the author's purpose in this respect has been fulfilled.

Many popular misconceptions to the contrary, neither the American Institute of Public Opinion nor the publisher, the Princeton University Press, is a part of Princeton University. Each, however, is bound to the university

by many ties. Although the AIPO is a privately owned commercial organization, its director is listed in the Princeton University catalogue as a member of the Advisory Councils of the Psychology Department and of the School of Public and International Affairs. The University's Office of Public Opinion Research serves as an official depository and reference guide for the end products of all AIPO, or Gallup, surveys—these archiving activities being under the direction of Princeton's Professor Hadley Cantril, and financed by the Rockefeller Foundation. In addition, the book, *Gauging Public Opinion*, by Cantril and various of his graduate students, which was first published in 1944 by the Princeton University Press, carries a formal dedication to George Gallup. The situation is, therefore, understandably confused; and one is often at a loss to know just where Gallup and his AIPO leave off and where Princeton University begins. However, despite this confusion, if *A Guide to Public Opinion Polls* is properly regarded simply as a popular publication by a business owner-manager about one of his several enterprises then it is subject to fewer strictures than could be levelled against it if it were otherwise considered. There is, after all, little justification for criticizing a book for not being scholarly or objective when it was never intended by the author that it contain these qualities.

The book, having been published prior to last November's presidential election, presents answers to the author's self-posed questions in a most assured, if not supercilious, fashion. The curious reader will find no mention of an "infant science" which latterly has become so popular a catchword with the commercial pollsters and their apologists. Critics of Gallup's procedures, when mentioned at all, are dismissed as being irresponsible, uninformed or undemocratic. Their criticisms are rationalized away either by retreat to a reverent quotation from James Bryce or by some breezy *non sequitur*. For example, Kornhauser's (not mentioned by name) careful exposition and criticism of the antilabor bias of AIPO from 1940 through 1945 is met by this reply: "Carried to the extreme, it would have meant that every time survey results showed the public hostile to a Hitler, another survey had to be found which was favorable to him."

Gallup gives an earnest defense of his "quota" method of sampling as compared to "area" or "probability" sampling. He also gives much lip service to statistical theory and Bernouillian "laws of probability" in discussing applicable formulae for determining size of sample in relation to probable error; and he compliments Theodore Brown and Samuel S. Wilks for the help they have given pollsters by "their studies on the statistical problems involved in sampling." The lay reader might well gain the impression that a considerable body of statistical theory underlies Gallup's "quota" operations, but the inquiring statistician will be relieved to find this admission on page 30: "While the accuracy of quota sampling cannot be determined from mathematical formulae—principally because of the inability to calculate the interviewer selection factor—there does exist a growing record of performance of public opinion polls which have in the past been operating on quota sampling procedures."

An entirely new section is devoted to reprinting a paper entitled "The Quintamensional Plan of Question Design." This was first delivered by Gallup in London before a meeting of the executives of his foreign affiliated enterprises, and later published in the *Public Opinion Quarterly*. The space consumed by this is one-thirteenth of the entire text of the book. One wonders at its disproportionate inclusion among "questions people most frequently ask about public opinion polls." It is difficult to imagine a deluge of letters and telegrams inquiring: "What is the quintamensional plan of question design?" More pertinent, however, would be an answer to the question concerning what proportion of AIPO surveys on controversial issues ever make use of this five-point plan involving the use of (*a*) filter or information questions, (*b*) open or free answer questions, (*c*) dichotomous or specific issue questions, (*d*) reason why questions, and (*e*) questions bearing upon intensity of opinion. The book gives us no clue as to the extent to which this laudable plan is actually used. On the surface it would seem too expensive for more than token use in a commercial operation.

A considerable compendium might be composed of important questions regarding opinion polling which are neither posed nor answered in this book. Some of these unanswered questions have become far more pressing than heretofore in view of the rent through the clouds which occurred last November, and many are currently being answered. But an extended discussion of these matters is not particularly apposite in a review.

The title is as misleading as that given to the first Kinsey report. Not only is this not a guide to opinion polls; it is not even a guide to all of Gallup's own polling activities—under whatever corporate names. Unfortunately for such organizations as Chicago's NORC, Michigan's Survey Research Center, and Columbia's Bureau of Applied Social Research, the public is all too likely to take the title at its face value.

If you want to learn in the shortest possible time the most complimentary things that Gallup has to say about the operations of his American Institute of Public Opinion, then this is the book for you.

The Decomposition of a Series of Observations Composed of a Trend, a Periodic Movement, and a Stochastic Variable. *A. Hald* (Assistant Professor, University of Copenhagen, Copenhagen, Denmark). Copenhagen, Denmark: G. E. C. Gads Forlag, 1948. Pp. 134. Paper.

REVIEW BY

D. B. DeLury, *Director, Department of Mathematical Statistics Ontario Research Foundation, Toronto 5, Canada*

AND

Boyd Harshbarger, *Professor of Statistics Virginia Polytechnic Institute, Blacksburg, Virginia*

THE TITLE of this work indicates clearly its content. The development concerns a series of observations $y_{\nu i}(\nu=1, 2 \cdots n; i=1, 2 \cdots k)$, with measurements of a concomitant variable $x_{\nu i}$ and assumes an underlying

model of the form

$$y_{\nu i} = \Gamma(x_{\nu i}) + \eta_i + \epsilon_{i\alpha},$$

where Γ denotes a trend, η a periodic element and ϵ a random component. The trend is assumed to be representable by a polynomial.

The fitting of this model to data is accomplished by constructing orthogonal polynomials in x (for the case in which the recorded x-values are in arithmetic progression and the y's have equal weights), such that, after elimination of the η's, the normal equations have diagonal form. Thus, the trend may be fitted one term at a time until a "satisfactory" fit is reached.

These orthogonal polynomials are, of necessity, considerably more complicated than those appropriate to fitting a trend only. Each pair of (n, k) values requires it own set of polynomials. Hence a table of values of these polynomials must be much more elaborate than that given, for example, in Fisher and Yates' tables. Part III of this study contains tabulated values of the polynomials up to degree 5 for the following (n, k) pairs.

$k = 3$	$n = 5, 7$
$k = 4$	$n = 25$
$k = 5$	$n = 5, 7, 8, 9, 10, 11$
$k = 12$	$n = 10$

From the several fields in which this mathematical model might conceivably be appropriate, the author selects two for special mention, the treatment of economic time series and the analysis of agricultural field experiments. The decision on the usefulness of the model in dealing with economic series must be left to economists. Its value in the design and analysis of field experiments will surely be widely questioned.

The agricultural experiment chosen to illustrate an application of these methods is called a "row experiment," in which the plots are placed consecutively in a row and divided into blocks, so that each treatment occurs once within each block. The arrangement of treatments within blocks is the same for all blocks. The x's then refer to the positions of the plots and the η's are treatment parameters.

One could, as the author remarks, employ this model when the treatments are arranged in any manner whatever, but when the "periodic" arrangement is abandoned, orthogonal polynomials are no longer feasible. In this case, regression analysis (presumably using ordinary orthogonal polynomials) or, what amounts to the same thing, a covariance analysis with the position of the plot as concomitant variable, permits the use of the same mathematical model. Even when the treatments are randomly arranged within blocks, control of the error resulting from variation in fertility is possible by these means and such procedures have, in fact, been used a number of times. It is true that the computational labor is considerably greater when orthogonal

polynomials, of the sort developed in this book, are not available, but then, when random arrangements are used, there is good reason to expect that such refinements will not be needed and that a simple analysis of variance will suffice. On the other hand, if one adopts a periodic arrangement of treatments in the hope that fertility variations are expressible in terms of a polynomial of low degree, his computational program is dismal indeed if this hope is not fulfilled. Furthermore, it is not clear to the reviewer that randomization can be dispensed with without assurance that variation in fertility is completely accounted for by the fitted polynomial (a viewpoint probably held also by the author, see p. 91), but presumably it never is.

Whether or not this mathematical model is suitable for routine use in field experiments, it seems likely that there are many situations in which it is quite appropriate and a table of orthogonal polynomials to facilitate the computations it entails is a useful item of statistical equipment. It is to be hoped that a more complete table than the one included in this book will soon be issued.

Part I, devoted to theory, includes a survey of standard regression theory and applies these classical methods to the problem at hand. The requisite orthogonal polynomials are developed and the necessary distribution theory is derived. The method of moving averages is discussed with respect to data to which this mathematical model is applicable and some comparison is made between this method, the type of regression analysis here developed and the analysis of variance as applied, for example, to a randomized block experiment. The methods used in this section are direct and the exposition is clear and concise.

Part II presents three worked examples, the fitting of a time series, the analysis of a row experiment and a re-working of the data of an industrial example reported by H. E. Daniels.*

The trend in the time series is fitted to degree 6 without reaching a good fit. The polynomial representing fertility variation in the row experiment is terminated at the fourth degree, on the ground that the coefficient of the fifth degree polynomial is not significant—a questionable criterion for general use. The same reason is given for choosing a third degree polynomial in the industrial example. Regression analysis in this last case comes out at about the same place as does the analysis of variance used by Daniels. Some advantage is claimed for the regression approach, in that only 3 degrees of freedom are needed to accomplish what 24 are used for in the analysis of variance.

Part II contains also a short but excellent discussion of some of the theoretical considerations on which experimental design is based.

This work is well put together and written in a concise and straightforward style that makes for easy reading. It can be recommended to those who (*a*) would like to have a concise, accurate statement of the methods and

* "Some Problems of Statistical Interest in Wool Research." *J Royal Stat Soc* 5:89–112 '38.

principal results of regression theory; (*b*) are interested in the theory of moving averages; (*c*) may find appropriate the mathematical model to which the book is devoted (e.g., in engineering research); (*d*) are perplexed about the old controversy over systematic and random arrangements, even though the concept of randomness is not dealt with as discerningly as one might wish.

Quality Control Methods. *Clifford W. Kennedy* (Quality Control Engineer, Federal Products Co., 1144 Eddy St., Providence, R. I.). New York 5: Prentice-Hall, Inc. (70 Fifth Ave.), 1948. Pp. vii, 243. $4.75, trade edition; $3 55, text edition. *Two reviews follow:*

REVIEW BY SEBASTIAN B. LITTAUER
Associate Professor of Industrial Engineering
Columbia University, New York 27, N. Y.

FOLLOWING the extensive as well as intensive application of statistical quality control during the war, the appearance of a number of texts at a variety of levels of interest and with a diversity of emphasis was to be expected. Of the dozen or so of books which have now been published, *Quality Control Methods* seems to be the first which is admittedly addressed to shop operating personnel, and therefore is to be viewed in that light. One might, on that account, reasonably have anticipated a greater emphasis on operating procedures, practical examples, administrative and economic problems, and in general the important nonstatistical aspects of quality control which have yet to be adequately presented. The fact, however, is that this work is devoted primarily to reporting statistical aspects of quality control, supposedly to fill the gaps left by other authors' omission of "basic details and primary concepts."

The role of statistics in quality control is introduced by a general discussion of sampling and the presentation of a variety of sampling plans including some dozen pages on sequential sampling. Following a discussion of "batch control" in Part 2, some elements of statistics are taken up in Part 3. Control charts for variables are considered in Part 4, and the closing section is devoted to administrative problems.

Statistical quality control can be motivated naturally by introducing the principles of acceptance sampling which can, again, lead naturally into control chart procedures. This has been effectively done by another author. In the present work, however, acceptance sampling procedures are presented apparently because "the progressive business house establishes a receiving system designed to make *certain* that the goods delivered fulfill requirements before the invoice is paid." The presentation of the various sampling plans is not prefaced by any exposition of statistical principles nor is their practical use validated by bringing in the role of statistical control in acceptance sampling practices. The exposition is necessarily descriptive and follows from

one plan to another, single, double and sequential, with little foundation that can be called statistical. It seems questionable, at best, to devote so much space to an attempted simplified explanation which "follows closely the instructions in *Sequential Analysis of Statistical Data: Applications*" when the readily available S.R.G. work does the job so well and so thoroughly.

The present treatment neglects to show the natural place of acceptance sampling in statistical quality control viewed both as a system of thought and as a system of practices. The author seems to confuse estimation with testing hypotheses. For example, early in Part 1, page 12, the author advises: "Throughout the use of sampling methods, never lose sight of the fact that basically you are attempting to estimate or judge the whole by what you see and think from examining a part or a sample." While this is altogether wholesome advice it does not apply to the use of the sampling plans presented, and, in fact, the originators of these plans would argue quite strongly against this practice. In a number of other places the author makes reference to estimating either the "condition" of the lot, or the quality of the lot.

Many statements made by the author must be read with great care in order to avoid misinterpretation. On reading (p. 7) that, "A sample is defined by the dictionary as a part shown to prove the quality of the whole," this reviewer checked some six dictionaries. The closest approach to the quoted statement was found to be, "A part of anything presented as evidence of the quality of the whole," from *Webster's Collegiate Dictionary* (Merriam and Co., 1917). Perhaps a statistical dictionary is needed. Again (p. 31), "The sample size n may be a number of ounces or pounds, a proportion of a fifty-ton gondola of ore, for example, or it may be so many drops of liquid, or 100 persons interviewed during a public opinion poll in a certain city" does not say what it may have been intended to mean. Nor can the statement (p. 68) "In other words, the chances would be 95 per cent good (with the buyers' risk from error in sampling, p, at 5 per cent) that the lot was of quality equal to $p_1(p_1=.05)$, or better" be regarded as correctly interpreting the use of a Wald sequential plan. In numerous other places in Part 1 the text requires discriminating reading that it is not likely to get from shop personnel who have had no previous statistical training. The whole treatment of acceptance sampling might have been clarified and simplified by use of OC curves. As it stands, the exposition is misleading to the uninitiated.

The remainder of the book is conventional except for the use of shop language. In the author's attempt to by-pass the preciseness of statistical inference for the suggestiveness of intuitive insight he has inadvertently exposed his whole exposition to question. Thus there is never a clear distinction between parameters and statistics, which leads to confusion between control limits and confidence intervals in the discussion of p-charts. This may also account for the absence of a clear-cut explanation of and emphasis

upon the concept of statistical control, stressing the importance of time order and rational subgrouping. There are no exercises for the student to work on and few adequate illustrative examples. In spite of the six practical problems in the last chapter, one might have hoped that this author would have embellished his writing with more material taken from his extensive experience.

It is rather surprising to this reviewer that there is no reference in the body of the text or in the index to Shewhart's contributions, and only a belated reference to his earlier book in the brief bibliography. In spite of all the space (26 pages) devoted to sequential sampling, Wald's name does not appear anywhere. On page 102, there is a replica of a table from page 23 of *Introduction to Industrial Statistics and Quality Control* by Paul Peach, with no specific acknowledgment. It is customary to acknowledge any complete reproduction apart from a general reference in the bibliography. Is it not the responsibility of the publisher to supply careful editing in order to pick up omissions like those mentioned.

REVIEW BY CHARLES R. SCOTT, JR.
Superintendent, Plant No. 1, SKF Industries, Inc.
Front St. and Erie Ave., Philadelphia 34, Pa.

THIS is an informative book about industrial quality control. The author handles the subject in a way that a beginner can understand. Rather than attempting to write a handbook putting forth a series of tables to be used in a mechanical sort of way, the author attempts to enable the beginner to secure a reasonable grasp of the statistical theory underlying industrial quality control. A discussion of the methods used in modern industrial plants carries the reader along through the theoretical sections of the book. The text is directed toward the practical man interested in statistical application of quality control at the operating level.

The book is divided into five sections. Part I discusses acceptance sampling, running through the gamut of sampling tables, sampling size, and the various accepted sampling methods. Part II discusses batch control and the practical methods of securing control by attribute inspection. Part III discusses and outlines practical uses of the frequency distribution and the standard deviation. In Part IV, Kennedy discusses average and range and compares the standard deviation with the average-range method. Of great importance is the way the author explains how to read a chart and interpret the data collected. Part V very nicely winds up with some good commonsense suggestions on what to do with the knowledge gained from the book. The reader is not left confused as to how to apply practically what he learned. The commercial examples form a basis for practical application. The book includes a bibliography, an appendix, and an index.

Metoder att Uppskatta Noggrannheten vid Linje- och Provytetaxering. [Methods of Estimating the Accuracy of Line and Sample Plot Surveys.] *Bertil Matérn* (Student in the Institute of Mathematical Statistics, Stockholm, and Research Statistician in the State Forest Research Institute, Stockholm). Stockholm, Sweden: State Forest Research Institute, 1947. Pp. 138. Paper.

Review by T. W. Anderson
Assistant Professor of Mathematical Statistics
Columbia University, New York 27, N. Y.

This book consists of a study of the use of information in a survey of a certain type for the estimation of the accuracy of that survey. This kind of systematic sample survey of a region is used, for example, in estimating certain characteristics of forests. The author utilizes the theory of stochastic processes in his treatment of the use of quadratic forms in observed variables for the estimation of the variances of estimates of various characteristics. The mathematical results are applied to a variety of problems arising in forestry surveys.

The kind of survey to which most attention is paid is a completely systematic sample. A specified grid of survey lines, say q, is placed in a specified way on the region Q, which is to be surveyed. The average of the characteristic on the grid lines, say $f(q)$, is the estimate of $f(Q)$, the average of the characteristic over the region. For instance, the estimate of the proportion of the area of Q which is forested is the proportion of the lengths of the survey lines running through forests. It is customary to estimate the "variance" of the error, $f(q) - f(Q)$, by means of a quadratic form in the values of the characteristic on pieces of the survey lines. The principal topic of this investigation is the selection of a quadratic form which will give a good estimate of the variance of the error regardless of the topographical variation.

The author suggests that a stationary stochastic process defined over a plane be used as a probability model for this problem. To each point (u, v) of the plane we attach a random variable $f(u, v)$ such that the set of random variables has the properties that the expectation of $f(u, v)$ is m and the variance is σ^2 and that the correlation of $f(u_1, v_1)$ and $f(u_2, v_2)$ is $\rho(t)$, where t is the distance between (u_1, v_1) and (u_2, v_2). If $\rho(t)$ is assumed continuous, we can define a random variable attached to a region as the integral of $f(u, v)$ over the region (with convergence meaning limit in the mean) divided by the area of the region. For example, $f(Q)$ is the integral of $f(u, v)$ over Q divided by the area of Q. The random variable associated with a line is a line integral of $f(u, v)$ divided by the length of the line. The expected value of a certain kind of quadratic form in such variables can be written as

$$\sigma^2 \int_0^t \rho(t)a(t)dt,$$

where $a(t)$ depends on the regions or lines involved.

The accuracy of a survey is indicated by the expected value $E\{T\}$, where $T = L[f(q) - f(Q)]^2$ and L is the length of q. The author finds an approximation to $E\{T\}$ and an associated "distance function" $\bar{a}(t)$. The expected values of different quadratic forms used to estimate $E\{T\}$ are studied by comparing the associated distance functions with $\bar{a}(t)$. Certain quadratic forms are suggested as best because their expected values are near $E\{T\}$ for all $\rho(t)$ in a certain class.

The remainder of the book deals with some questions of the relation of the model to reality and with the application of the theory to practical problems. Plot surveys, sampling trees on the lines and double-sampling schemes are also considered. A number of examples are given, and computational procedures are occasionally suggested. Although the main text is in Swedish, there is an extensive summary in English (20 pages), and the table and figure headings are given in both languages.

This study is of interest to statisticians on two counts, firstly as a study of stochastic processes from the point of view of statistical application and secondly as a presentation of means of choosing methods of estimating the accuracy of surveys. Of course, between the theory and the application is the interpretation of the probability model. A line survey here is a completely systematic survey; there is no chance mechanism involved in placing the lines. To enable him to apply probability calculus the author considers the formation of the topography under the grid lines as a stochastic process. The combined effects of Nature and man in giving the region the characteristics to be surveyed is taken to be that of a process defined by this model. Such an interpretation of the model leads to results different from those that would be obtained by considering placing a grid at random on a region with a given topography.

Except for one feature of this study, this approach seems to give results that have reasonable interpretations for the practical problems. In fact, answers are given to questions which have previously been answered inadequately or not at all. In devoting the remainder of the review to the one questionable feature the reviewer does not wish to detract from the otherwise fine work.

The purpose of a survey q is to estimate some characteristic of a given region Q. One would like to say something about the discrepancy between the estimate, say $f(q)$, and the value for this specific region, say $f(Q)$. For example, if $f(q)$ were normally distributed with mean value $f(Q)$ and variance τ^2, one could state the inequality $f(q) - \tau t \leq f(Q) \leq f(q) + \tau t$ with confidence α if t were chosen so the integral of the standard normal density from $-t$ to t were α. Confidence α would be interpreted as being approximately the proportion of times such a statement would be correct if the statement were made for a number of situations (specified by $f(q_1)$, $f(Q_1)$, τ_1^2; $f(q_2)$, $f(Q_2)$, τ_2^2; . . .). α would be independent of how $f(Q)$ would be chosen (as long as the choice is independent of $f(q)$. In particular, if $f(Q)$ were fixed the above theory would hold.

In his treatment the author takes $f(Q)$ to be a random variable, not a parameter. Thus, $E\{[f(q)-f(Q)]^2\}$ is an average over all possible topographical variations defined by the stochastic process over Q. But this is not the average which one would like to use to define the accuracy of a survey of a specific area. One would like the average taken over all topographical variations for which $f(Q)$ is the value that this region actually has. In short, one would like $f(Q)$ to be treated as a constant.

The probability model specifies a distribution of the set $f(u, v)$ which is the product of the distribution of $f(Q)$ and the conditional distribution of the set $f(u, v)$ given $f(Q)$. The use of the distribution of $f(Q)$ implies that we believe that $f(Q)$ for the particular region Q we survey arose in a way described by the model. Now this is unrealistic because we may have chosen Q to survey simply because we think $f(Q)$ is in a certain range of values. This question of using Bayes Theorem has been discussed so much that there is no need of doing more in a review than pointing out that this is what the author does.

It is possible to avoid this difficulty by using the conditional distribution of $f(u, v)$ given $f(Q)$. Then the question arises as to whether the mathematical results of this investigation hold for the conditional distribution. Since the author's conclusions are based on the distribution of $f(q_j)$ (values on line segments q_j) and $f(Q)$, we need only consider the conditional joint distribution of $f(q_j)$ given $f(Q)$. The author's conclusions rest on approximate equalities of the sort $E\{T_k\}=E\{T\}$, where T_k is a quadratic form in $f(q_j)$. Do these hold if $f(Q)$ is held fixed; that is, is $E\{T_k|f(Q)\}=E\{T|f(Q)\}$ true? It can be shown that in general the second equality does not follow from the first. This is easily demonstrated if the original process is normal because then the joint distribution of $f(q_j)$ and $f(Q)$ is normal. Thus, the actual conclusion we wish to draw from the theoretical study does not follow directly.

It is possible, however, that these results hold in some approximate sense. If the region Q is large and if $\rho(t)$ decreases rapidly, $f(Q)$ is distributed with a small variance, and, hence, may behave nearly like a constant. Then the effect of holding $f(Q)$ fixed may not be great. However, in order to justify his conclusions the author needs to study this question.

The author has exercised considerable ingenuity in solving difficult mathematical problems. The development is rigorous, except that many approximations are used; the effects of the approximations are studied to verify that they are legitimate. The problems of application considered seem to be important and interesting. It is to be hoped that the author in a future study will show that the objection to the use of Bayes Theorem approach can be met.

Lecture Notes on Mathematical Theory of Probability and Statistics. *Richard von Mises* (Gordon McKay Professor of Aerodynamics and Applied Mathematics, Harvard University, Cambridge, Mass.). Graduate School of Engineering, Special Publication No. 1. Cambridge 38, Mass.: Harvard Graduate School of Engineering (c/o Miss Natalie Nicholson, Librarian, Pierce Hall), 1947. Pp. 300. Paper, mimeographed. Out of print.

REVIEW BY BENJAMIN EPSTEIN
Associate Professor of Mathematics, Wayne University
Detroit 1, Michigan

THESE are notes on a year course in probability and statistics given for undergraduates and graduates at Harvard University. The author has, in the opinion of this reviewer, given an excellent presentation of some of the most fundamental ideas in probability and statistics. It is a pity that these notes are not generally available. They would represent a welcome addition in a field where the really good elementary texts can be counted on the fingers of one hand.

The frequency concept of probability is the dominant theme throughout the text and the author gives a very clear introduction to the basic notions underlying his approach to the axiomatic foundations of probability. In particular he defines at the outset the concepts of limiting frequency, chance as distinguished from randomness, collective, and the basic operations of place selection, mixing, partition, and combination.

There are many workers in probability and statistics who object to the way in which von Mises formulates his axioms, to the way in which formal and empirical aspects of probability are mixed.* Personally this reviewer prefers the treatment of statistics as given by Cramér in his recent book, *Mathematical Methods in Statistics.* However, I am of the opinion that it would be a mistake to make the preference so strong that one becomes blind to the originality and the basic character of the von Mises approach. While there appear to be grounds for preferring the measure-theoretic approach (which is of course at heart also a frequency theory) to that of von Mises, one should recall that von Mises was a pioneer in the field of probability and that he grappled, as far back as 1919, with many of the problems underlying the foundations of probability.

This reviewer, when reading the text, was made keenly aware more than once of the fact that the author is a master in the field and is a man who is not satisfied with a superficial glossing over of difficulties. He meets difficulties head on and brings them out into the light of day where they can be analyzed. This is particularly true in Chapters 3, 6, and 9. In these chapters he deals with (*a*) the weak law of large numbers; (*b*) Bayes' theorem and its relation to the problem of inference; (*c*) certain aspects of the Neyman-

* For an evaluation of this point, *see:*

1. Mises, R. von. "On the Foundations of Probability and Statistics." *Ann Math Stat* 12:191–205 Je '41.
2. Doob, J. L. "Probability as Measure." *Ann Math Stat* 12:206–14, Je '41.
3. Mises, R. von, and Doob, J. L. "Discussion of Papers on Probability Theory." *Ann Math Stat* 12:215–17 Je '41.

Pearson theory of testing hypotheses; (*d*) the contrast between Bayes' theorem and the Neyman-Pearson theory of testing hypotheses; and (*e*) the confidence interval concept. He shows very clearly the role played by the a priori distribution of the unknown parameter or parameters in determining the kind of probability inferences that one can draw from a sample. These are precisely the sort of things which should be brought to the attention of the student who is just entering the field of statistics.

The short sections on runs and Mendelian heredity are also very good and illustrate important statistical concepts. Another valuable feature of the book is that many of the problems given to the student are non-trivial and thought-provoking. The problems are drawn from many fields and help to illuminate parts of the theory that might otherwise have remained unclear to the student.

The treatment throughout is rigorous and the author does not hesitate to call on techniques in modern analysis which are about on the level of those given in Whittaker and Watson's *Modern Analysis*. Theorems are carefully stated and where concepts are involved the author takes the trouble to point out what would happen if certain underlying assumptions were not satisfied.

The selection of certain sections or chapters for special comment is not to be taken to mean that the rest of the book is not well presented. The reviewer has only picked out a few of the things that were treated *unusually well* in this book and that are either omitted or glossed over in most texts. Some of the other topics are taken up as well in a book such as Cramér's.

By way of summary, it is my opinion that the continual emphasis by von Mises on underlying concepts makes good sense scientifically and pedagogically. I also feel that this book gives the student some insight into the historical development of the subject and into the nature of the early difficulties and paradoxes. It is particularly the beginning student who needs to be reminded over and over again of just what assumptions are made when one makes a probability statement or an inference from data. It is refreshing to find an author who is honest enough to make the following explicit statement (Chap. 6, p. 16): "It remains an invariable fact, dominating all problems in mathematical statistics, that *no substantial inference can be drawn from a small number of observations if nothing is known 'a priori,' i.e., preliminary to the experiments, about the object of experimentation.*"

Statistics in School. *W. L. Sumner* (Senior Lecturer in Education, University College, Nottingham, England). Oxford, England: Basil Blackwell & Mott Ltd. (49 Broad St.), 1948. Pp. vii, 183. 9*s.* 6*d.*

REVIEW BY F. G. CORNELL
Professor of Education and Director
Bureau of Research and Service, College of Education
University of Illinois, Champaign, Illinois

THIS book, evidently intended for use as a text, is a summary of lectures given to postgraduate students in education at the University College, Nottingham. As such, it lacks some of the features which might be expected

of a book purposely prepared as a textbook. The author recognizes the difficulties of students of education with standard mathematical treatments on statistics. This is a problem of teacher training institutions not only in Great Britain, but also in the United States. Unfortunately, a successful formula has not yet appeared for the intuitive, genuinely nonmathematical, yet technically adequate and sufficiently comprehensive, treatment of statistics which would reach larger numbers of educational leaders and the rank and file of teachers.

The author of this book is reasonably successful in some of his attempts "to make it as simple as possible." Included are some topics needed and usually not found in educational statistics such as an introduction to analysis of variance and factorial analysis. With these, however, there are questions of the value of the limited depth of treatment to which the author restricts himself evidently to avoid complexity. To what extent can a "simplified" text compromise with the criterion of adequacy of content coverage, for instance, by including factorial analysis but limiting its discussion to tetrad differences? In his presentation of partial correlation, the author covers only the three-variable problem without multiple regression, error of estimate, and other concepts useful in understanding the subject.

The chapter on correlation and regression is typical of what is likely to result from confining the subject of educational statistics to 183 pages. In that chapter there appear technical errors of omission or commission which would lead an immature student to incomplete or incorrect information. Regression is introduced primarily as a means of defining the Pearson-r. Thus, only the deviation form of the regression equation is given. The essential concepts of "error of estimate" and "prediction" oddly appear under a section headed *Spearman's Footrule*, instead of in the section on "regression" where it appears to belong. In this chapter, some technical and sematic difficulties appear in such condensed statements as: "In (c) there is perfect correspondence between the scores and correlation is complete *and there is no regression.*" In education and psychology we usually find that correlation, if present, is *partial positive correlation.*" Incidentally in a later chapter the author gives the old probable error for r formula, $.6745\ (1-r^2)/\sqrt{N}$, and the Fisher z transformation as "hyperbolic arctangent," in a way which would not encourage use of the latter.

Certain important omissions prevent this book from being a significant addition to the field. Not included is a modern approach to sampling. A careful distinction is not made between large sample and small sample method, "statistic" and "parameter." Notions of sampling distribution are introduced through the medium of the "probable error" anachronism. Chi square is presented with the limited emphasis of "goodness of fit" and contingency.

The book throughout would be improved if there was greater use made of applications to educational problems and if examples were not excluded more or less to the rather sterile subject of school marks. The writing is heavily orientated in the direction of marking, to which relatively little

concern has been given in education in this country in recent years. A broader orientation of a book on school statistics seems essential if it is to contribute significantly to the study of modern educational problems.

Mathematical Treatment of the Results of Agricultural and Other Experiments, Second Edition. *M. J. van Uven* (Professor in the Agricultural University of Wageningen, Netherlands). Groningen-Batavia, Netherlands: P. Noordhoff F.V., 1946. Pp. viii, 310.

REVIEW BY G. A. BAKER
Assistant Professor of Mathematics and Assistant Statistician in the Experiment Station, College of Agriculture University of California, Davis, California

THE PRESENT tendency in the design and analysis of field trials seems to be to carry the elaboration of the mathematical models first devised by R. A. Fisher to greater and greater extremes of complication. The theory of many such models is well worked out and tables provided so that the probabilities of the errors in decision made in any trial can be estimated. The difficulty is that the Fisherian models, in many cases, do not sufficiently approach reality in their fundamental assumptions about the behavior of soil fertility. Professor van Uven is refreshingly careful and thorough in his fundamental assumptions about the behavior of soil fertility. He assumes, that soil fertility is a continuous function of position that changes slowly. His methods in this respect agree with Neyman's method of parabolic curves (1929). There is some evidence that, in some cases at least, this assumption of slowness of change in soil fertility is not sufficiently realistic.

Professor van Uven's tests of significance are based on estimates of mean differences and standard errors obtained by the methods of least-squares and asymptotic averaging. Reference is then made to the normal probability curve regardless of the sample size and character of the population distribution. The whole discussion of significance is naive as compared with those of J. Neyman, E. S. Pearson, and R. A. Fisher.

Professor van Uven discusses some of the simpler Fisherian designs in detail. In this respect he says that although sometimes the influences at work on the yield will not fully justify the hypotheses yet in many cases they will be very well applicable. This remark, of course, applies to situations with which Professor van Uven is familiar.

A list of technical terms in nine languages is included.

The second edition is a photographic reproduction of the first edition published in 1935. Professor van Uven intended to make some changes based on his further experience in teaching and examining field trials but could not because of shortages due to the war.

The book is well and logically written and proceeds on the basis of assumptions that are clearly stated. It will well repay study by anyone interested in the analysis of field trial data.

The Fourier Transforms of Probability Distributions. *Aurel Wintner* (Professor of Mathematics, Johns Hopkins University, Baltimore, Md.). Baltimore 18, Md.: the Author (Rowland Hall, Charles and 34th Sts.), 1947. Pp. v, 185. Paper. $3.00.

REVIEW BY J. WOLFOWITZ
Associate Professor of Mathematical Statistics
Columbia University, New York 27, N. Y.

THIS publication is a set of notes, taken by Dr. F. W. Light, of a course given by Professor Wintner in 1942–3. It begins with a description of a (one-dimensional) distribution function. Successive chapters are entitled: transforms, moments, convolutions, convergent convolutions, convergence in measure, semi-groups (i.e., infinitely divisible laws), inversions, examples, and projections (i.e., multidimensional and marginal distributions). Some references are given at the end of the book.

The reviewer immensely enjoyed reading the book. Since it is a set of notes it does not treat the subject exhaustively and gives us only glimpses of the author's attractive mathematical style. Yet it proceeds pleasantly and easily, and readily absorbs the reader's attention.

The reviewer would, however, like to take exception to the choice of subjects and certain aspects of their presentation. In spite of its name, this is not a book on Fourier transforms; it is a book on distribution functions studied with the aid of Fourier transforms. For example, the question of when a function is a characteristic function is barely discussed. To the best of the reviewer's knowledge the phrase "chance variable" or its equivalent is never mentioned. Yet the author is in effect compelled to introduce chance variables to clarify certain concepts. His section on "Examples" discusses mappings which are in effect chance variables of only moderate interest for the student of probability theory. There seems little point to so scrupulous an avoidance of the customary terminology of probability theory, with its attendant values of brevity and suggestiveness. All the "chance variables" which occur in this book are independent; this is often an unnecessary limitation. The question of the convergence of a series of independent chance variables is given considerable attention, yet neither the law of large numbers nor the central limit theorem are even mentioned.

The reviewer was disappointed by the extreme paucity of appropriate references. The zero or one law, which is due to Kolmogoroff (generalization by Paul Lévy) is wrongly ascribed by the author to Borel; the reference is *Rendiconti Palermo*, Vol. 29, 1909, where a special case of the strong law of large numbers is proved. The following minor remarks occur to the reviewer: (1) The proof (p. 168), that the distribution of Student's ratio is the same for all underlying joint distributions which are radially symmetric, can be considerably shortened. It suffices to note that Student's ratio is homogeneous of degree zero; the statement is really valid for all such functions, provided only that the functions are undefined on at most a set of probability

zero. (2) A simple proof by means of Fourier transforms of the fact that the totality of marginal distributions of every linear combination of k chance variables uniquely determines the joint distribution of the k variables (p. 160) is due to Cramér and Wold and to be found in the former's book, *Random Variables and Probability Distributions.*

The reviewer is of the opinion that, while no advanced worker in mathematics will want to be without a copy of this book, it is not particularly suited for statisticians or elementary students of probability theory. This is a great pity, and it is to be hoped that the distinguished author can be persuaded to treat the subject more comprehensively and systematically in a future book.

Random Normal Deviates: 25,000 Items Compiled From Tract No. XXIV (M. G. Kendall and B. Babington Smith's Tables of Random Sampling Numbers). *Herman Wold* (Professor of Statistics and Director of the Institute of Statistics, University of Uppsala, Uppsala, Sweden). University of London, University College, Department of Statistics, Tracts for Computers. London N.W.1: Cambridge University Press (Bentley House, 200 Euston Road), 1948. Pp. xiii, 51. Paper. 5*s*.

REVIEW BY H. BURKE HORTON
Bureau of Transport Economics and Statistics
Interstate Commerce Commission, Washington, D. C.

STATISTICIANS are indebted to Professor Wold for substantially increasing the number of random normal deviates available for general use. Prior to the publication of this set of 25,000 items there were in published form only 10,400 such items compiled by P. C. Mahalanobis from L. H. C. Tippett's random numbers.

In illustrating statistical processes a relatively small quantity of random normal deviates will usually suffice. However, for certain important research purposes, such as experimental deduction of the distribution function of a sampling statistic, large quantities of deviates are required. In this booklet Professor Wold has provided statisticians with 25,000 random normal deviates, each recorded to two decimal places. The items were derived from the Kendall-Smith tables of random numbers, column by column. It is unfortunate that Professor Wold's labor was doubled, due to the fact that use of the Kendall-Smith tables by rows yielded a set for which the variance was too large by a significant amount ($P = 0.7\%$). In the words of the author, "It is difficult to decide whether the failure with the first set [*by rows*] is accidental or due to some slight defect in the construction of Kendall-Smith's tables." If the latter possibility is the source of trouble, such difficulties may in the future be avoided, or at least minimized, by the use of recently developed convergent processes to generate the underlying set of random digits.

For the convenience of the user, sums and sums of squares for the fifty

items of each column are included. Tests for local randomness based upon sums, sums of squares, ranges, and sign runs, were applied to each page (500 items), to blocks of 5,000 items, and to the entire set. The results were in accord with the hypothesis of normality. Subsets yielding unusual test results are listed for the benefit of users of small sets from the table. For clarity, in line 2, page vii, substitute ".0100 and .9900" for "0 and 1".

Textual material accompanying the tables is clearly and concisely written. The author presents illustrated techniques for the construction of univariate, bivariate, and multivariate normal distributions, with specified underlying parameters. The material presented on the construction of a multivariate normal distribution is of particular value as a reference. This booklet will be a useful addition to the libraries of statisticians and statistical organizations requiring sizeable quantities of random normal deviates for research or educational purposes.

Say It With Figures. *Hans Zeisel* (McCann-Erickson, New York City). New York 16: Harper & Brothers (383 Madison Ave.), 1947. Pp. xix, 250. $3.00.

Review by Gregor Sebba
Professor of Economics, The University of Georgia
Athens, Georgia

The calamity that recently befell the public opinion polls points up the fact that the development of statistical methodology and techniques has by far outpaced the analysis of what Professor Paul F. Lazarsfeld, in his introduction to Hans Zeisel's book *Say It With Figures* aptly terms "the conceptional meaning of statistical procedures." Since Albert B. Blankenship reviewed the book in this Journal (42: 666–7 D '47) from the point of view of commercial research only, it seems appropriate to discuss briefly its contribution to conceptual analysis and its usefulness to teachers of statistics.

Zeisel's book deals with three broad subjects: "Problems of Classification" (Part I), "Means of Numerical Presentation" (Part II) and "Tools of Causal Analysis" (Part III). Part I is primarily meant for users of the questionnaire and interview methods; Zeisel's discussion of "Don't know" answers (which can easily be adapted to "Undecided" answers) is particularly illuminating when applied to political opinion polls. But it is Parts II and III upon which the importance of the book rests. Among the "Means of Numerical Presentation," Dr. Zeisel singles out percentages and simple indices for a penetrating analysis of their logic. The use of per cent figures is not generally advisable but needs "specific justification" and can be decided upon "only with a complete background of concrete data and specific circumstances" (p. 72). Their use for comparing increase or decrease in two or more populations, in particular, is logically justified only if the change is (or is treated as being) "in exact proportion to the factors chosen as a base for per cent computation" (p. 80). It thus turns out that per cent comparisons

"offer only approximate solutions" since they merely discount "*a priori* the effects of concomitant variates" (R. A. Fisher); hence their use for comparison "will be justified [only] to the extent to which this a priori reasoning proves correct" (p. 81). In a two-dimensional table, per cents should be run in the direction of the variable to be studied for its effect, provided the sample is representative in this direction; if it is not, proper weighting becomes necessary (pp. 105–6). Of particular interest is Zeisel's subsequent study of the problem of reducing three- and more-dimensional tables (chap. 6), a problem arising because "only tables containing two variables can be presented in their entirety and still be clearly readable" (p. 127). There follows an illuminating discussion of simple indices of the type developed in sociometrics, leading up to the warning that "there is a certain danger that somewhere along the way from a clearly defined object to its mathematical symbols, the clarity of thought is lost; indices sometimes pretend to measure a concept which . . . turns out to be ambiguous and, therefore, not measurable. Neither a descriptive label nor an impressive mathematical formula are a safeguard against . . . indices which do not measure what they purport to measure" (p. 166).

Part III, "Tools of Causal Analysis," contains a superior treatment of cross-tabulation as a tool of research. Cross-tabulation "refines" and "explains"—though the explanation may turn out to be spurious; the distinction between "true" and "spurious" inter-variable correlation depending on whether or not the correlation reflects a direct causal connection, i.e., whether the explaining factor is asymmetrically or symmetrically connected with the two variables (p. 202). Answering the question when to cross-tabulate, the author sets down the rule that if a result is analyzed successively by various breakdowns and it is known or suspected that some of them are interrelated, then these interrelated breakdowns should be tabulated, not successively, but simultaneously (p. 203). A critical discussion of the panel technique of interviewing concludes the book.

Dr. Zeisel's study represents a step forward on the road indicated by such classical earlier treatises as Zizek's study of averages (1913), Winkler's monograph on relatives (1923) and Haberler's analysis of index numbers (1927). The book might be termed an essay in the logic of statistics procedures. The teacher of elementary and applied statistics will find in it useful numerical examples and charts of great forcefulness; and while much of the author's discussion is too refined for beginners, there remains enough to enable the teacher to go through the initial chapters of an elementary text without putting the students (and himself) to sleep.

Although Professor Lazarsfeld claims that "some of the inevitable errata" have been corrected in the 1948 printing, others seem to have passed unnoticed; among them an ugly blemish: in Table X-19 (pp. 240–2), there occur three questions and answers of the form: "If Hitler offered peace now . . . , would you favor or oppose such a peace?" Answer: "Yes."

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44 SEPTEMBER 1949 NUMBER 247

ARTICLES

BOOK REVIEWS

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 247 SEPTEMBER 1949 *Volume 44*

THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM

Los Alamos Laboratory

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

ALREADY in the nineteenth century a sharp distinction began to appear between two different mathematical methods of treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of *sets of particles*. In pure mathematics an intensive study of the properties of sets of points was the subject of a new field. This is the so-called theory of sets, the basic theory of integration, and the twentieth century development of the theory of probabilities prepared the formal apparatus for the use of such models in theoretical physics, i.e., description of properties of aggregates of points rather than of individual points and their coordinates.

Soon after the development of the calculus, the mathematical apparatus of partial differential equations was used for dealing with the problems of the physics of the continuum. Hydrodynamics is the most widely known field formulated in this fashion. A little later came the treatment of the problems of heat conduction and still later the field theories, like the electromagnetic theory of Maxwell. All this is very well known. It is of course important to remember that the study of the

physics of the continuum was paralleled through "kinetic theories." These consist in approximating the continuum by very large, but finite, numbers of interacting particles.

When a physical problem involves an intermediate situation, i.e., a system with a moderate number of parts, neither of the two approaches is very practical. The methods of analytical mechanics do not even give a qualitative survey of the behavior of a system of three mutually attractive bodies. Obviously the statistical-mechanical approach would also be unrealistic.

An analogous situation exists in problems of combinatorial analysis and of the theory of probabilities. To calculate the probability of a successful outcome of a game of solitaire (we understand here only such games where skill plays no role) is a completely intractable task. On the other hand, the laws of large numbers and the asymptotic theorems of the theory of probabilities will not throw much light even on qualitative questions concerning such probabilities. Obviously the practical procedure is to produce a large number of examples of any given game and then to examine the relative proportion of successes. The "solitaire" is meant here merely as an illustration for the whole class of combinatorial problems occurring in both pure mathematics and the applied sciences. We can see at once that the estimate will never be confined within given limits with certainty, but only—if the number of trials is great—with great probability. Even to establish this much we must have recourse to the laws of large numbers and other results of the theory of probabilities.

Another case illustrating this situation is as follows: Consider the problem of evaluating the volume of a region in, say, a twenty-dimensional space. The region is defined by a set of inequalities

$$f_1(x_1, x_2 \cdots x_{20}) < 0; f_2(x_1, x_2 \cdots x_{20}) < 0; \cdots f_{20}(x_1, x_2 \cdots x_{20}) < 0.$$

This means that we consider all points$(x_1, x_2, x_3, \cdots x_{20})$ satisfying the given inequalities. Suppose further that we know that the region is located in the unit cube and we know that its volume is not vanishingly small in general. The multiple integrals will be hardly evaluable. The procedure based on the definition of a volume or the definition of an integral, i.e., the subdivision of the whole unit cube, for example, each coordinate x_1 into ten parts, leads to an examination of 10^{20} lattice points in the unit cube. It is obviously impossible to count all of them. Here again the more sensible approach would be to take, say 10^4 points

at random from this ensemble and examine those only; i.e., we should count how many of the selected points satisfy all the given inequalities. It follows from simple application of ergodic theorems that the estimate should be, *with great probability*, valid within a few per cent.

As another illustration, certain problems in the study of cosmic rays are of the following form. An incoming particle with great energy entering the atmosphere starts a whole chain of nuclear events. New particles are produced from the target nuclei, these in turn produce new reactions. This cascade process continues with more and more particles created until the available individual energies become too small to produce further nuclear events. The particles in question are protons, neutrons, electrons, gamma rays and mesons. The probability of producing a given particle with a given energy in any given collision is dependent on the energy of the incoming particle. A further complication is that there is a probability distribution for the direction of motions. Mathematically, this complicated process is an illustration of a so-called Markoff chain. The mathematical tool for the study of such chains is matrix theory. It is obvious that in order to obtain a mathematical analysis, one would have to multiply a large number of $(n \times n)$ matrices, where n is quite great.

Here again one might try to perform a finite number of "experiments" and obtain a class or sample of possible genealogies. These experiments will of course be performed not with any physical apparatus, but theoretically. If we assume that the probability of each possible event is given, we can then play a great number of games of chance, with chances corresponding to the assumed probability distributions. In this fashion one can study empirically the asymptotic properties of powers of matrices with positive coefficients, interpreted as transition probabilities.

II

Finally let us consider more generally the group of problems which gave rise to the development of the method to which this article is devoted. Imagine that we have a medium in which a nuclear particle is introduced, capable of producing other nuclear particles with a distribution of energy and direction of motion. Assume for simplicity that all particles are of the same nature. Their procreative powers depend, however, on their position in the medium and on their energy. The problem of the behavior of such a system is formulated by a set of integro-differential equations. Such equations are known in the kinetic theory of gases as the Boltzmann equations. In the theory of probabilities one

has somewhat similar situations described by the Fokker-Planck equations. A very simplified version of such a problem would lead to the equation:

$$\frac{\partial u(x, y, z)}{\partial t} \quad a(x, y, z)\Delta u + b(x, y, z)u(x, y, z) \tag{1}$$

where $u(x, y, z)$ represents the density of the particles at the point (x, y, z). The Laplacian term, $a\Delta u$ on the right hand side corresponds to the diffusion of the particles, and bu to the particle procreation, or multiplication. [In reality, the equation describing the physical situation stated above is much more complicated. It involves more independent variables, inasmuch as one is interested in the density $w(x, y, z; v_x, v_y, v_z)$ of particles in phase space, v being the velocity vector.] The classical methods for dealing with these equations are extremely laborious and incomplete in the sense that solutions in "closed form" are unobtainable. The idea of using a statistical approach at which we hinted in the preceding examples is sometimes referred to as the Monte Carlo method.

The mathematical description is the study of a flow which consists of a mixture of deterministic and stochastic processes.[1] It requires its own laws of large numbers and asymptotic theorems, the study of which has only begun. The computational procedure looks in practice as follows: we imagine that we have an ensemble of particles each represented by a set of numbers. These numbers specify the time, components of position and velocity vectors, also an index identifying the nature of the particle. With each of these sets of numbers, random processes are initiated which lead to the determination of a new set of values. There exists indeed a set of probability distributions for the new values of the parameters after a specified time interval Δt. Imagine that we draw at random and *independently*, values from a prepared collection possessing such distributions. Here a distinction must be made between those parameters which we believe vary independently of each other, and those values which are strictly determined by the values of other parameters. To illustrate this point: assume for instance that in the fission process the direction of the emitted neutron is independent of its velocity. Or again, the direction of a neutron in a homogeneous medium does not influence the distance between its origin and the site of its first collision. On the other hand, having "drawn" from appropri-

[1] von Neumann, J., and Ulam, S., *Bulletin A.M.S.*, Abstract 51-9-165 (1945).

ate distributions the velocity of a new-born particle and the distance to its first collision, the time elapsed in travel is completely determined and has to be calculated accordingly. By considering a large number of particles with their corresponding sets of parameters we obtain in this fashion another collection of particles and a new class of sets of values of their parameters. The hope is, of course, that in this manner we obtain a good sample of the distributions at the time $t+\Delta t$. This procedure is repeated as many times as required for the duration of the real process or else, in problems where we believe a stationary distribution exists, until our "experimental" distributions do not show significant changes from one step to the next.

The essential feature of the process is that we avoid dealing with multiple integrations or multiplications of the probability matrices, but instead sample single chains of events. We obtain a sample of the set of all such possible chains, and on it we can make a statistical study of both the genealogical properties and various distributions at a given time.

III

We want now to point out that modern computing machines are extremely well suited to perform the procedures described. In practice, the set of values of parameters characterizing a particle is represented, for example, by a set of numbers punched on a card. We have at the outset a large number of particles (or cards) with parameters reflecting given initial distributions. The step in time consists in the production of a new such set of cards. The original set is processed one by one by a computing machine somewhat as follows: The machine has been set up in advance with a particular sequence of prescribed operations. These divide roughly into two classes: (1) production of "random" values with their frequency distribution equal to those which govern the change of each parameter, (2) calculation of the values of those parameters which are deterministic, i.e., obtained algebraically from the others. It may seem strange that the machine can simulate the production of a series of random numbers, but this is indeed possible. In fact, it suffices to produce a sequence of numbers between 0 and 1 which have a uniform distribution in this interval but otherwise are uncorrelated, i.e., pairs will have uniform distribution in the unit square, triplets uniformly distributed in the unit cube, etc., as far as practically feasible. This can be achieved with errors as small as desired or practical. What is more, it is not necessary to store a collection of such numbers in the machine itself, but paradoxically enough the machine can

be made to produce numbers simulating the above properties by iterating a well-defined arithmetical operation.

Once a uniformly distributed random set is available, sets with a prescribed probability distribution $f(x)$ can be obtained from it by first drawing from a uniform uncorrelated distribution, and then using, instead of the number x which was drawn, another value $y=g(x)$ where $g(x)$ was computed in advance so that the values y possess the distribution $f(y)$.

Regarding the sequence of operations on a machine, more can be and has been done. The choice of the *kind* of step to be performed by the machine can be made to depend on the values of certain parameters just obtained. In this fashion even dependent probabilistic processes can be performed. Quite apart from mechanized computations, let us point out one feature of the method which makes it advantageous with, say, stepwise integration of differential equations. In order to find a particular solution, the usual method consists in iterating an algebraical step, which involves in the nth stage values obtained from the $(n-1)$th step. The procedure is thus serial, and in general one does not shorten the time required for a solution of the problem by the use of more than one computer. On the other hand, the statistical methods can be applied by many computers working in parallel and independently. Several such calculations have already been performed for problems of types discussed above.[2]

IV

Let us indicate now how other equations could be dealt with in a similar manner. The first, purely mathematical, step is to transform the given equation into an equivalent one, possessing the form of a diffusion equation with possible multiplication of the particles involved. For example as suggested by Fermi, the time-independent Schrödinger equation

$$\Delta\psi(x, y, z) = (E - V)\psi(x, y, z)$$

could be studied as follows. Re-introduce time dependence by considering

$$u(x, y, z, t) = \psi(x, y, z)e^{-Et}$$

u will obey the equation

$$\frac{\partial u}{\partial t} = \Delta u - Vu.$$

[2] Among others, problems of diffusion of neutrons, gamma rays, etc. To cite an example involving the study of matrices, there is a recent paper by Goldberger, *Phys. Rev.* 74, 1269 (1948), on the interaction of high energy neutrons with heavy nuclei.

This last equation can be interpreted however as describing the behavior of a system of particles each of which performs a random walk, i.e., diffuses isotropically and at the same time is subject to multiplication, which is determined by the value of the point function V. If the solution of the latter equation corresponds to a spatial mode multiplying exponentially in time, the examination of the spatial part will give the desired $\psi(x, y, z)$—corresponding to the lowest "eigenvalue" E.

The mathematical theory behind our computational method may be briefly sketched as follows: As mentioned above and indicated by the examples, the process is a combination of stochastic and deterministic flows.[1] In more technical terms, it consists of repeated applications of matrices—like in Markoff chains—and completely specified transformations, e.g., the transformation of phase space as given by the Hamilton differential equations.

One interesting feature of the method is that it allows one to obtain the values of certain given operators on functions obeying a differential equation, without the point-by-point knowledge of the functions which are solutions of the equation. Thus we can get directly the values of the first few moments of a distribution, or the first few coefficients in the expansion of a solution into, for example, a Fourier series without the necessity of first "obtaining" the function itself. "Symbolically" if one is interested in the value of $U(f)$ where U is a functional like the above, and f satisfies a certain operator equation $\psi(f)=0$, we can in many cases obtain an idea of the value of $U(f)$ directly, without "knowing" f at each point.

The asymptotic theorems so far established provide the analogues of the laws of large numbers, such as the generalizations of the weak and strong theorems of Bernoulli, Cantelli-Borel.[3] The more precise information corresponding to that given in the Laplace-Liapounoff theory of additive processes has not yet been obtained for our more general case. In particular it seems very difficult to estimate in a precise fashion the probability of the error due to the finiteness of the sample. This estimate would be of great practical importance, since it alone would allow us to suit the size of the sample to the desired accuracy.

The "space" in which our process takes place is the collection of all possible chains of events, or infinite branching graphs.[4] The general properties of such a phase space have been considered but much work remains to be done on the specific properties of such spaces, each corresponding to a given physical problem.

---

[3] Everett, C. J. and Ulam, S., U S.A.E.C., Los Alamos reports LADC-533 and LADC-534. Declassified, 1948.

[4] Everett, C. J. and Ulam, S., *Proc Nat. Acad. Sciences*, 34,403 (1948).

APPLICATIONS OF SOME SIGNIFICANCE TESTS FOR THE MEDIAN WHICH ARE VALID UNDER VERY GENERAL CONDITIONS*

JOHN E. WALSH
The RAND Corporation

In two other papers ([1] and [2]) order statistics were used to derive some tests for the population median which have significance levels which are either exact or bounded under some very general conditions. These order statistic tests were found to be very efficient for small samples from a normal population; also they can be applied with very little computation. This paper contains applications of these tests to several well known statistical problems. Also a graphical method of applying the tests of [1] is outlined.

INTRODUCTION

THE SIGNIFICANCE tests for the population median derived in [1] are valid if the n observations on which a test is based are independent and are drawn from n populations satisfying the conditions

(A)
1) Each population is continuous (i.e. its cumulative distribution is continuous).
2) Each population is symmetrical.
3) The median of each population has the same value ϕ.

These tests compare ϕ with a given constant value ϕ_0.

The tests for comparing ϕ with ϕ_0 derived in [2] are based on the assumption of a sample from a normal population. The significance level of these tests, however, is bounded near the value for normality if the n independent observations are from populations necessarily satisfying only conditions (A).

An important feature of conditions (A) is that no two of the observations are necessarily drawn from the same population. Thus the order statistic tests can be applied to a wide variety of situations.

The motivation for introducing tests with bounded significance levels in addition to the tests of [1] was to obtain a wider variety of suitable significance levels for small values of n without greatly weakening the generality of application.

* The results presented in this paper were obtained in the course of research conducted under the sponsorship of the Office of Naval Research. This research was performed while the author was at Princeton University.

TABLE 1

SOME ONE-SIDED AND SYMMETRICAL SIGNIFICANCE TESTS FOR $n \leq 15$

n	Significance Level of Tests: One sided	Significance Level of Tests: Symmetrical	Tests — *Symmetrical:* Accept $\phi \neq \phi_0$ if either — *One-sided:* Accept $\phi < \phi_0$ if	*One-sided:* Accept $\phi > \phi_0$ if	Approx. Efficiency for normality
4	6.2%	12.5%	$x_4 < \phi_0$	$x_1 > \phi_0$	95%
5	6.2%	12.5%	$\frac{1}{2}(x_4+x_5) < \phi_0$	$\frac{1}{2}(x_1+x_2) > \phi_0$	98%
	3.1%	6.2%	$x_5 < \phi_0$	$x_1 > \phi_0$	96%
6	4.7%	9.4%	$\max\,[x_5, \frac{1}{2}(x_4+x_6)] < \phi_0$	$\min\,[x_2, \frac{1}{2}(x_1+x_3)] > \phi_0$	97%
	3.1%	6.2%	$\frac{1}{2}(x_5+x_6) < \phi_0$	$\frac{1}{2}(x_1+x_2) > \phi_0$	98%
	1.6%	3.1%	$x_6 < \phi_0$	$x_1 > \phi_0$	95%
7	5.5%	10.9%	$\max\,[x_5, \frac{1}{2}(x_4+x_7)] < \phi_0$	$\min\,[x_3, \frac{1}{2}(x_1+x_4)] > \phi_0$	95%
	2.3%	4.7%	$\max\,[x_6, \frac{1}{2}(x_5+x_7)] < \phi_0$	$\min\,[x_2, \frac{1}{2}(x_1+x_3)] > \phi_0$	98%
	1.6%	3.1%	$\frac{1}{2}(x_6+x_7) < \phi_0$	$\frac{1}{2}(x_1+x_2) > \phi_0$	98%
	0.8%	1.6%	$x_7 < \phi_0$	$x_1 > \phi_0$	95%
8	4.3%	8.6%	$\max\,[x_6, \frac{1}{2}(x_4+x_8)] < \phi_0$	$\min\,[x_3, \frac{1}{2}(x_1+x_5)] > \phi_0$	94.5%
	2.7%	5.5%	$\max\,[x_6, \frac{1}{2}(x_5+x_8)] < \phi_0$	$\min\,[x_3, \frac{1}{2}(x_1+x_4)] > \phi_0$	96%
	1.2%	2.3%	$\max\,[x_7, \frac{1}{2}(x_6+x_8)] < \phi_0$	$\min\,[x_2, \frac{1}{2}(x_1+x_3)] > \phi_0$	98%
	0.8%	1.6%	$\frac{1}{2}(x_7+x_8) < \phi_0$	$\frac{1}{2}(x_1+x_2) > \phi_0$	98%
	0.4%	0.8%	$x_8 < \phi_0$	$x_1 > \phi_0$	95%
9	5.1%	10.2%	$\max\,[x_6, \frac{1}{2}(x_4+x_9)] < \phi_0$	$\min\,[x_4, \frac{1}{2}(x_1+x_6)] > \phi_0$	91%
	2.2%	4.3%	$\max\,[x_7, \frac{1}{2}(x_5+x_9)] < \phi_0$	$\min\,[x_3, \frac{1}{2}(x_1+x_5)] > \phi_0$	96%
	1.0%	2.0%	$\max\,[x_8, \frac{1}{2}(x_5+x_9)] < \phi_0$	$\min\,[x_2, \frac{1}{2}(x_1+x_5)] > \phi_0$	95.5%
	0.6%	1.2%	$\max\,[x_8, \frac{1}{2}(x_7+x_9)] < \phi_0$	$\min\,[x_2, \frac{1}{2}(x_1+x_3)] > \phi_0$	99%
	0.4%	0.8%	$\frac{1}{2}(x_8+x_9) < \phi_0$	$\frac{1}{2}(x_1+x_2) > \phi_0$	98%
10	5.6%	11.1%	$\max\,[x_6, \frac{1}{2}(x_4+x_{10})] < \phi_0$	$\min\,[x_5, \frac{1}{2}(x_1+x_7)] > \phi_0$	87.5%
	2.5%	5.1%	$\max\,[x_7, \frac{1}{2}(x_5+x_{10})] < \phi_0$	$\min\,[x_4, \frac{1}{2}(x_1+x_6)] > \phi_0$	93%
	1.1%	2.1%	$\max\,[x_8, \frac{1}{2}(x_6+x_{10})] < \phi_0$	$\min\,[x_3, \frac{1}{2}(x_1+x_5)] > \phi_0$	96.5%
	0.5%	1.0%	$\max\,[x_9, \frac{1}{2}(x_6+x_{10})] < \phi_0$	$\min\,[x_2, \frac{1}{2}(x_1+x_5)] > \phi_0$	96.5%
11	4.8%	9.7%	$\max\,[x_7, \frac{1}{2}(x_4+x_{11})] < \phi_0$	$\min\,[x_5, \frac{1}{2}(x_1+x_8)] > \phi_0$	
	2.8%	5.6%	$\max\,[x_7, \frac{1}{2}(x_5+x_{11})] < \phi_0$	$\min\,[x_5, \frac{1}{2}(x_1+x_7)] > \phi_0$	89%
	1.1%	2.1%	$\max\,[\frac{1}{2}(x_6+x_{11}), \frac{1}{2}(x_8+x_9)] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_6), \frac{1}{2}(x_3+x_4)] > \phi_0$	
	0.5%	1.1%	$\max\,[x_9, \frac{1}{2}(x_7+x_{11})] < \phi_0$	$\min\,[x_3, \frac{1}{2}(x_1+x_5)] > \phi_0$	97%
12	4.7%	9.4%	$\max\,[\frac{1}{2}(x_4+x_{12}), \frac{1}{2}(x_5+x_{11})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_9), \frac{1}{2}(x_2+x_8)] > \phi_0$	
	2.4%	4.8%	$\max\,[x_8, \frac{1}{2}(x_5+x_{12})] < \phi_0$	$\min\,[x_5, \frac{1}{2}(x_1+x_8)] > \phi_0$	
	1.0%	2.0%	$\max\,[x_9, \frac{1}{2}(x_6+x_{12})] < \phi_0$	$\min\,[x_4, \frac{1}{2}(x_1+x_7)] > \phi_0$	93.5%
	0.5%	1.1%	$\max\,[\frac{1}{2}(x_7+x_{12}), \frac{1}{2}(x_9+x_{10})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_6), \frac{1}{2}(x_3+x_4)] > \phi_0$	
13	4.7%	9.4%	$\max\,[\frac{1}{2}(x_4+x_{13}), \frac{1}{2}(x_5+x_{12})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)] > \phi_0$	
	2.3%	4.7%	$\max\,[\frac{1}{2}(x_5+x_{13}), \frac{1}{2}(x_6+x_{12})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_9), \frac{1}{2}(x_2+x_8)] > \phi_0$	
	1.0%	2.0%	$\max\,[\frac{1}{2}(x_6+x_{13}), \frac{1}{2}(x_9+x_{10})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_8), \frac{1}{2}(x_4+x_5)] > \phi_0$	
	0.5%	1.0%	$\max\,[x_{10}, \frac{1}{2}(x_7+x_{13})] < \phi_0$	$\min\,[x_4, \frac{1}{2}(x_1+x_7)] > \phi_0$	94.5%
14	4.7%	9.4%	$\max\,[\frac{1}{2}(x_4+x_{14}), \frac{1}{2}(x_5+x_{13})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_{11}), \frac{1}{2}(x_2+x_{10})] > \phi_0$	
	2.3%	4.7%	$\max\,[\frac{1}{2}(x_5+x_{14}), \frac{1}{2}(x_6+x_{13})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_2+x_9)] > \phi_0$	
	1.0%	2.0%	$\max\,[x_{10}, \frac{1}{2}(x_6+x_{14})] < \phi_0$	$\min\,[x_5, \frac{1}{2}(x_1+x_9)] > \phi_0$	90.5%
	0.5%	1.0%	$\max\,[\frac{1}{2}(x_7+x_{14}), \frac{1}{2}(x_{10}+x_{11})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_8), \frac{1}{2}(x_4+x_5)] > \phi_0$	
15	4.7%	9.4%	$\max\,[\frac{1}{2}(x_4+x_{15}), \frac{1}{2}(x_5+x_{14})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_{12}), \frac{1}{2}(x_2+x_{11})] > \phi_0$	
	2.3%	4.7%	$\max\,[\frac{1}{2}(x_5+x_{15}), \frac{1}{2}(x_6+x_{14})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_{11}), \frac{1}{2}(x_2+x_{10})] > \phi_0$	
	1.0%	2.0%	$\max\,[\frac{1}{2}(x_6+x_{15}), \frac{1}{2}(x_{10}+x_{11})] < \phi_0$	$\min\,[\frac{1}{2}(x_1+x_{10}), \frac{1}{2}(x_5+x_6)] > \phi_0$	
	0.5%	1.0%	$\max\,[x_{11}, \frac{1}{2}(x_7+x_{15})] < \phi_0$	$\min\,[x_5, \frac{1}{2}(x_1+x_9)] > \phi_0$	92%

The purpose of this paper is to use these two types of tests to obtain generalized solutions for the cases of quality control, slippage, and the sign test. Also direct applications of the tests are considered.

Tables 1 and 2 contain a list of some practically important one-sided and symmetrical tests of the two types considered ($x_1, \cdots, x_n$ represent the values of the n observations arranged in increasing order of magnitude). The tests of Tables 1 and 2 were chosen so that the significance levels are approximately 5%, 2.5%, 1%, 0.5% for one-sided tests and 10%, 5%, 2%, 1% for symmetrical tests; also so that the amount of computation required for the application of a test is small.

To clarify the use of the tests of Table 1, consider the following example: Let $n=10$ and the values of the 10 observations be

$$0.7,\ -1.1,\ -0.2,\ -1.2,\ 0.1,\ 3.4,\ 3.7,\ 1.3,\ 1.8,\ 2.0.$$

The hypothesis to be tested is that the common median of the continuous symmetrical populations from which these independent observations were drawn has the value zero. Then $\phi_0=0$ and

$$\begin{array}{ll} x_1 = -1.2 & x_6 = 1.3 \\ x_2 = -1.1 & x_7 = 1.8 \\ x_3 = -0.2 & x_8 = 2.0 \\ x_4 = 0.1 & x_9 = 3.4 \\ x_5 = 0.7 & x_{10} = 3.7 \end{array}$$

Apply the symmetrical test of Table 1 at the 5.1% significance level to these observations. Then

$$\max\ [x_7,\ \tfrac{1}{2}(x_5 + x_{10})] = \max\ (1.8,\ 2.2) = 2.2 > 0$$
$$\min\ [x_4,\ \tfrac{1}{2}(x_1 + x_6)] = \min\ (0.1,\ 0.1) = 0.1 > 0.$$

Hence ϕ is significantly different from zero at the 5.1% significance level.

As an example of the application of the tests of Table 2 let $n=6$ and the observations be

$$-0.01,\ 0.21,\ -0.72,\ 0.00,\ -0.05,\ -1.81.$$

The hypothesis to be tested is that the common median of the continuous symmetrical populations from which these six independent observations were drawn has the value 0.1. Then $\phi_0=0.1$ and

$$\begin{array}{ll} x_1 = -1.81 & x_4 = -0.01 \\ x_2 = -0.72 & x_5 = 0.00 \\ x_3 = -0.05 & x_6 = 0.21. \end{array}$$

TABLE 2

SOME ONE-SIDED AND SYMMETRICAL TESTS WITH BOUNDED SIGNIFICANCE LEVELS

Tests	*SYMMETRICAL:* Accept $\phi \neq \phi_0$ if either		Significance Level for Normality		Significance Level Bounds for Conditions (A)				Approx. Efficiency for normality
	ONE-SIDED: Accept $\phi < \phi_0$ if	*ONE-SIDED:* Accept $\phi > \phi_0$ if			One-sided Tests		Symmetrical Tests		
n			One-sided	Symmetrical	Upper	Lower	Upper	Lower	
4	$1.055x_4 - .055x_1 < \phi_0$	$1.055x_1 - .055x_4 > \phi_0$	5%	10%	6.2%		12.5%		96%
5	$.63x_5 + .37x_4 < \phi_0$ $1.02x_5 - .02x_1 < \phi_0$	$.63x_1 + .37x_2 > \phi_0$ $1.02x_1 - .02x_5 > \phi_0$	5% 2.5%	10% 5%	6.2% 3.1%	3.1%	12.5% 6.2%	6.2%	99% 97%
6	$.63x_6 + .37x_5 < \phi_0$ $1.06x_6 - .06x_1 < \phi_0$	$.63x_1 + .37x_2 > \phi_0$ $1.06x_1 - .06x_6 > \phi_0$	2.5% 1%	5% 2%	3.1% 1.6%	1.6%	6.2% 3.1%	3.1%	98.5% 98%
7	$.785x_7 + .215x_6 < \phi_0$ $1.05x_7 - .05x_1 < \phi_0$	$.785x_1 + .215x_2 > \phi_0$ $1.05x_1 - .05x_7 > \phi_0$	1% 0.5%	2% 1%	1.6% 0.8%	0.8%	3.1% 1.6%	1.6%	97% 96%
8	$\max[x_7, (.5x_8 + .28x_6 + .22x_7)] < \phi_0$ $.785x_8 + .215x_7 < \phi_0$	$\min[x_2, (.5x_1 + .28x_3 + .22x_2)] > \phi_0$ $.785x_1 + .215x_2 > \phi_0$	approx. 1% 0.5%	approx. 2% 1%	1.2% 0.8%	0.8% 0.4%	2.3% 1.6%	1.6% 0.8%	98% 97%
9	$\max[x_8, (.5x_9 + .28x_7 + .22x_8)] < \phi_0$	$\min[x_2, (.5x_1 + .28x_3 + .22x_2)] > \phi_0$	approx. 0.5%	approx. 1%	0.6%	0.4%	1.2%	0.8%	98.5%

Apply the symmetrical test of Table 2 which has a 5% significance level if the n observations are a sample from a normal population. Then

$$.63x_6 + .37x_5 = 0.132 + 0.000 = .132 > 0.1$$

$$.63x_1 + .37x_2 = -1.146 - 0.266 = -1.412 < 0.1.$$

Hence the median is not significantly different from 0.1 on the basis of this test. If the six observations are a sample from a normal population, the significance level of the test is 5%. If only conditions (A) are necessarily satisfied, however, the significance level of the test is bounded between 6.2% and 3.1%.

As pointed out in [1], if a symmetrical population has a mean, the mean equals the median. Thus the order statistic tests are tests of the mean if the populations considered have a mean and conditions (A) are satisfied.

The efficiencies listed for the tests in Tables 1 and 2 refer to the power efficiency of the order statistic test considered for the case in which the n observations are a sample from a normal population. The definition of the power efficiency of a test is given in section 3 of [1]. Essentially determination of the power efficiency of a test consists in finding the sample size (not necessarily integral) of the most powerful test of the specified hypothesis (in this case the t-test) which has approximately the same power function as the given test; this sample size divided by the sample size for the given test is called the power efficiency of that test. An intuitive explanation of the meaning of power efficiency is given in [3]. Roughly speaking, the power efficiency of a test is the percentage of the total available information per observation which is being utilized by that test.

Section 2 contains an outline of a graphical method of applying certain of the tests derived in [1]. The remaining sections contain methods of applying the tests listed in Tables 1 and 2 to several well known statistical problems. For simplicity the tests of Tables 1 and 2 will be referred to as the tests of section 1. If tests for $n>15$ or near different significance levels than those approximated by the tests of Tables 1 and 2 are required, such tests can usually be obtained from the results given in [1].

GRAPHICAL APPLICATION OF TESTS

Let us consider a graphical method applying the tests of Table 1. Since max $(x, y)<\phi_0$ has the interpretation that both x and y are less than ϕ_0 while min $(x, y)>\phi_0$ means that both x and y are greater than ϕ_0, it is only necessary to develop a method of deciding how expressions

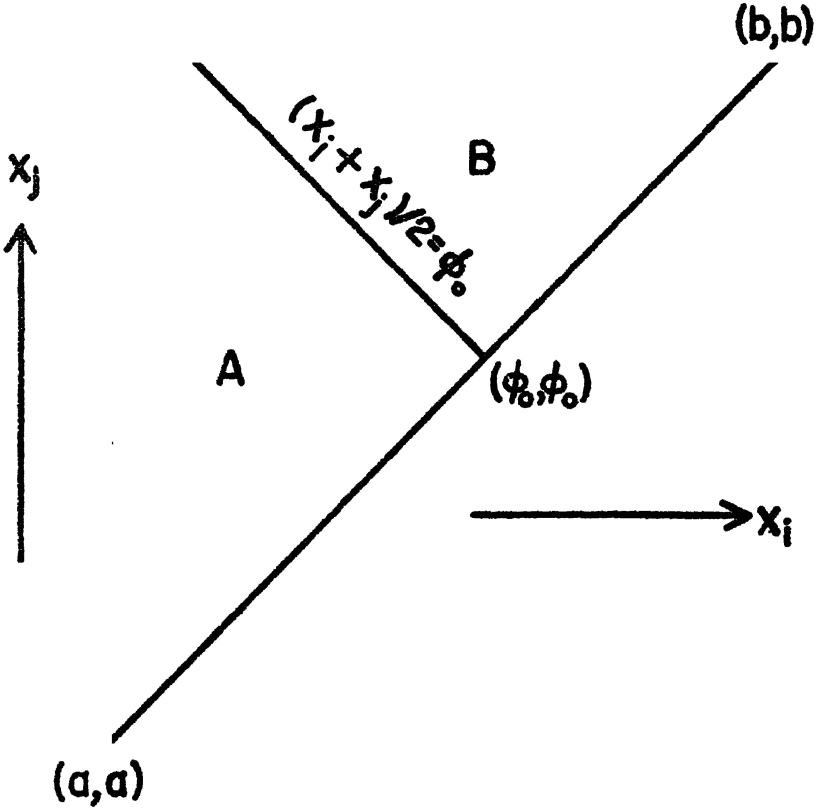


FIG. 1. SCHEMATIC DIAGRAM OF REGIONS A AND B

of the forms x_i and $\frac{1}{2}(x_i+x_j)$ compare with ϕ_0. Direct comparison accomplishes this for the case of x_i. Thus it is sufficient to determine a graphical method of deciding whether $\frac{1}{2}(x_i+x_j)>\phi_0$ or $\frac{1}{2}(x_i+x_j)<\phi_0$.

For nearly all situations to which the tests of Table 1 would be applied, the n observations can be considered to have practical upper and lower limits, say a and b. Using these limits a graphical method of deciding when $\frac{1}{2}(x_i+x_j)<\phi_0$, $(i<j)$, consists in constructing the region A of the x_i, x_j-plane defined by

$$\tfrac{1}{2}(x_i + x_j) < \phi_0, \qquad x_i < x_j, \qquad a \leqq x_i, x_j \leqq b.$$

If the point (x_i, x_j) falls in this region, $\frac{1}{2}(x_i+x_j)<\phi_0$. Similarly $\frac{1}{2}(x_i+x_j)>\phi_0$ if (x_i, x_j) falls into the region B defined by

$$\tfrac{1}{2}(x_i + x_j) > \phi_0, \qquad x_i < x_j, \qquad a \leqq x_i, x_j \leqq b.$$

Fig. 1 contains a schematic diagram of the regions A and B. These regions are particularly easy to construct because the line $\frac{1}{2}(x_i+x_j)=\phi_0$ is perpendicular to the line $x_i=x_j$ at the point (ϕ_0, ϕ_0).

As the regions do not depend on i or j, a single graph will suffice for all the one-sided or symmetrical tests of Table 1. (It is assumed that the bounds a and b are the same for each test.)

DIRECT APPLICATIONS

One important application of the tests of section 1 consists in using these tests as substitutes for the corresponding t-tests. The order statistic tests are more easily applied, valid under more general conditions, and approximately as efficient as the corresponding t-tests.

A second application occurs in cases where it is reasonably certain that the observations are from populations satisfying conditions (A) but there is no reason to suppose that the observations have the same precision. As an example, consider the examination of a gravimetric method of determining the amount of calcium oxide in given samples whose CaO content is known (see [4]). The results for 10 given samples are as follows:

CaO Present	CaO Found By Method	Ratio
Mg.	Mg.	
4.0	3.7	.925
8.0	7.8	.975
12.5	12.1	.968
16.0	15.6	.975
20.0	19.8	.990
25.0	24.5	.980
31.0	31.1	1.003
36.0	35.5	.986
40.0	39.4	.985
40.0	39.5	.988

If it can be assumed that the method used to determine the amount of CaO is symmetrical, the above ratios satisfy conditions (A) with mean (median) equal to unity when the null hypothesis that the average amount of CaO found by the given method equals the true amount holds. Since the amount of CaO present varies from 4.0 mg. to 40.0 mg., the populations from which the ratios are considered drawn may have variances which differ noticeably. Thus application of the t-test to these 10 ratios is a questionable procedure. As conditions (A) are satisfied, however, the tests of section 1 are directly applicable. Apply the symmetrical test of Table 1 at the 1.0% significance level to these 10 ratios. Then $\phi_0=1$ and

$$\max\,[x_9,\ \tfrac{1}{2}(x_6+x_{10})] = \max\,(0.990,\ 0.994) = 0.994 < 1.000$$

$$\min\,[x_2,\ \tfrac{1}{2}(x_1+x_5)] = \min\,(0.968,\ 0.950) = 0.950 < 1.000.$$

Thus the method examined yields an average value which is significantly different from the true value at the 1.0% significance level.

USE OF TRANSFORMATIONS

Consider a situation where the n populations from which the n observations were drawn are continuous, have the same median, but are not symmetrical. (A sample from any continuous non-symmetrical population satisfies these conditions.) In practical cases it is sometimes known that replacing each observation value x by the value $g(x)$, where $g(y)$ is a continuous strictly monotonically increasing function of y, will result in a set of observations from approximately symmetrical populations. If ϕ is the common population median for the original observations, $g(\phi)$ will be the population median for the transformed observations. Thus the transformed observations are from populations approximately satisfying conditions (A) with median $g(\phi)$.

Now $g(x_i)$ will be the i^{th} largest of the values of the transformed observations if x_i is the i^{th} largest of the original observations. Also $g(\phi)>g(\phi_0)$ if and only if $\phi>\phi_0$. Similarly for $g(\phi)<g(\phi_0)$ and $g(\phi)\neq g(\phi_0)$. Thus the tests of section 1 are easily modified to obtain tests of $\phi<\phi_0$, $\phi>\phi_0$, and $\phi\neq\phi_0$. The procedure followed is to replace x_i by $g(x_i)$ and ϕ_0 by $g(\phi_0)$ in the body of Tables 1 and 2 but leave everything else as is. For example, for $n=9$ the one-sided test

Accept $\phi<\phi_0$ *if* $\max[x_8,\ \tfrac{1}{2}(x_7+x_9)]<\phi_0$.

is modified to the test.

Accept $\phi<\phi_0$ *if* $\max\ \{g(x_8),\ \tfrac{1}{2}[g(x_7)+g(x_9)]\}<g(\phi_0)$.

As another example let $n=8$. Then the one-sided test

Accept $\phi>\phi_0$ *if* $\min\ [x_2,\ (.5x_1+.28x_3+.22x_2)]>\phi_0$.

is modified to the test

Accept $\phi>\phi_0$ if $\min\ \{g(x_2),\ [.5g(x_1)+.28g(x_3)+.22g(x_2)]\}>g(\phi_0)$.

The choice of the function $g(y)$ will usually depend on past experience with the type of situation being investigated. For example, in some cases replacing each observation value by the log of that value has been found to yield observations from approximately symmetrical populations.

Since only symmetry is required, there may exist many suitable transformations which have not been used in the past because they do not yield observations from approximately normal populations.

It should be emphasized that tests obtained by transformations are not necessarily also tests of the means of the original populations.

QUALITY CONTROL APPLICATIONS

The tests of section 1 are readily adapted to control chart use. In addition to being valid under extremely general conditions, very efficient for normality, and easily applied, the resulting control chart tests for the median have the valuable property of being independent of the dispersion control chart tests in the sense that the construction and application of the median control charts do not depend on any dispersion values.

Two quality control situations are considered: Control with respect to a given standard value of the median; control with no standard given.

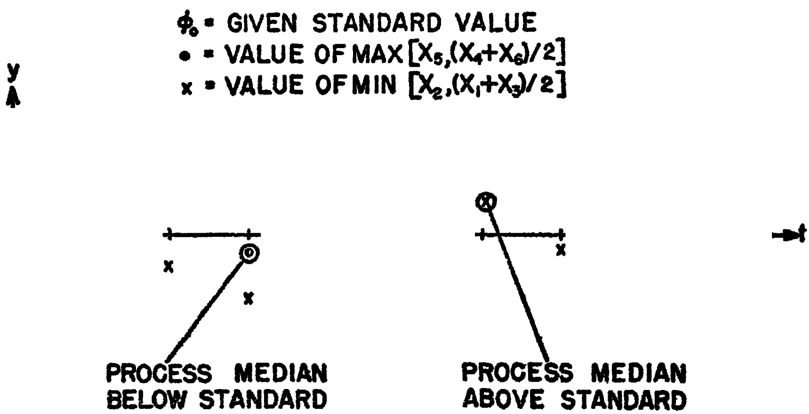


FIG. 2. CONTROL CHART OF MEDIAN FOR SETS OF OBSERVATIONS OF SIZE 6.

As pointed out in section 1, if a symmetrical population has a mean, the mean equals the median. Thus if each observation is drawn from a population satisfying the additional condition that the mean exists, the following quality control tests for the median are also tests for the mean.

A. *Control with respect to given standard.* In this case a standard value ϕ_0 of the median ϕ is given. The method used in constructing and applying control charts based on section 1 tests is demonstrated by the following example: Consider construction of a control chart based on sets of observations of size 6. The time t is plotted on the horizontal axis of the chart while a function y of the six observations is plotted on the vertical axis. The central line on the chart is $y=\phi_0$. The function $y=\max\ [x_5,\ \frac{1}{2}(x_4+x_6)]$ is plotted on the chart with dots while the function $y=\min\ [x_2,\ \frac{1}{2}(x_1+x_3)]$ is plotted with crosses. A dot falling below

the central line indicates that the true value of the median is below the standard value; a cross above the central line indicates that the true value of the median is above the standard value. Fig. 2 furnishes an example of how this control chart might look in application. If conditions (A) hold and the true value of the median equals the given standard value, the probability of a dot falling below the central line equals 4.7%. Similarly for a cross falling above the line. The probability of either a dot falling below the central line or a cross appearing above it is 9.4%. Thus, if the observations are from populations satisfying conditions (A), the control chart exemplified by Fig. 2 represents a graphical method of continually applying a symmetrical test of $\phi \neq \phi_0$ at the 9.4% significance level. If a one-sided test of $\phi < \phi_0$ at the 4.7% significance level is all that is desired, plot only the values of max $[x_5, \frac{1}{2}(x_4+x_6)]$ on the chart. If a one-sided test of $\phi > \phi_0$ at the 4.7% significance level is sought, plot only the values of min $[x_2, \frac{1}{2}(x_1+x_3)]$.

The method of constructing control charts outlined above is directly applicable to all the section 1 tests. As another example, consider sets of size 5. Let a dot be defined as $1.02x_5 - .02x_1$ and a cross by $1.02x_1 - .02x_5$. Then Fig. 2 represents a continual application of a symmetrical test of $\phi \neq \phi_0$ with significance level 5% for normality and upper bound 6.2% for conditions (A). Plotting $1.02x_1 - .02x_5$ alone furnishes a one-sided test of $\phi > \phi_0$ with significance level 2.5% for normality and upper bound 3.1% for conditions (A), etc.

Control charts based on unequal size sets of observations can also be readily obtained by use of the tests of section 1. The method used to obtain control charts for these cases consists in giving a separate definition of what a dot and cross are to represent for each set size. The significance level can differ according to size or be approximately the same for all sizes. As an example of the procedure used, consider a control

TABLE 3

DEFINITION OF DOTS AND CROSSES

Set Size	Definition of Dot	Definition of Cross
5	$1.02x_5 - .02x_1$	$1.02x_1 - .02x_5$
6	$.63x_6 + .37x_5$	$.63x_1 + .37x_2$
7	max $[x_6, (x_5+x_7)/2]$	min $[x_2, (x_1+x_3)/2]$
8	max $[x_6, (x_5+x_8)/2]$	min $[x_3, (x_1+x_4)/2]$
9	max $[x_7, (x_5+x_9)/2]$	min $[x_3, (x_1+x_5)/2]$
10	max $[x_7, (x_5+x_{10})/2]$	min $[x_4, (x_1+x_6)/2]$

chart using sets varying from 5 to 10 observations in size. Let the dots and crosses for Fig. 2 be defined by Table 3. Then, if conditions (A) hold, Fig. 2 represents a continual application of a symmetrical test of $\phi \neq \phi_0$ at approximately the 5% significance level. The dots alone furnish a one-sided test of $\phi < \phi_0$ at approximately the 2.5% level while the crosses by themselves represent a one-sided $\phi > \phi_0$ at approximately the 2.5% significance level.

B. *Control—no standard given* (*control chart test*). For this case a standard value of the median is not given. The standard value is replaced by an estimate of the true median value made on the basis of past data taken while the process was in control. Once a suitable estimate of the median is obtained, the determination of control charts for this case becomes identical with that considered in section A if ϕ_0 is replaced by the estimated value of the median. Hence the main problem is to obtain a suitable estimate for the median on the basis of the given past observations.

A very satisfactory estimate can be obtained if the past observations were drawn from populations satisfying conditions (A) and the additional condition that the first three moments of each population are finite. Then the average of all the past observations furnishes an estimate of the true value of the median. This estimate has the favorable properties:

1. The expected value of the estimate equals the true value of the median.
2. The estimate tends to the true median value as the number of past observations increases.

From an application viewpoint, the additional condition that the first three moments of each population are finite is not very restrictive; this condition is satisfied for nearly all populations arising in practice.

Use of the above estimate in place of ϕ_0 allows the control chart methods developed in section A to be utilized.

GENERALIZED SLIPPAGE TEST

The usual slippage situation investigated is the following: A change is made with affects a continuous population in such a way that the shape of the population distribution remains fixed but the population mean may move. This slippage of the mean is tested on the basis of samples drawn from the population both before and after the change.

The purpose of this section is to generalize the above slippage problem and present a solution to the generalized situation. The generalized slippage problem is the following: A change is made which affects m

continuous populations; k of these populations, ($k=0, 1, \cdots, m$), are affected in such a way that each population has the same shape distribution before and after the change; the remaining $m-k$ populations are symmetrical both before and after the change (but the distribution shapes may change). It is required to test whether the change affected the values of the means of the m populations.

The method used to derive this test consists in obtaining n dummy observations, ($n \geqq 6$), which satisfy conditions (A) with zero median when the null hypothesis that all the means remained fixed (no slippage) is true. Slippage tests with a wide variety of significance levels can then be obtained by applying tests based on conditions (A) to these dummy observations. In particular, if $n \leqq 15$, the tests of Tables 1 and 2 of section 1 can be used.

The procedure used to obtain this test is the following:

(a) $1 \leqq m \leqq 5$. Choose r such that $6 \leqq rm \leqq 15$. Then draw r samples of size s from each population both before and after the change. Record the order in which these samples were drawn. Form the mean of each sample. Consider the mean of the i^{th} sample drawn from the j^{th} population, ($i=1, \cdots, r; j=1, \cdots, m$), after the change. Subtract this mean from the mean of the i^{th} sample drawn from that population before the change. Under the conditions of the generalized slippage problem it is easily seen that the resulting rm dummy observations satisfy conditions (A) with zero median if there is no slippage.

(b) $m \geqq 6$. Draw a sample of size s from each population both before and after the change. Form the mean of each sample. Subtract the mean of the sample drawn from the j^{th} population after the change, ($j=1, \cdots, m$), from the mean of the sample drawn from that population before the change. The resulting m dummy observations satisfy conditions (A) with zero median if the null hypothesis of no slippage is true.

In both cases (a) and (b) a set of independent dummy observations satisfying conditions (A) with zero median are obtained under the null hypothesis of no slippage. The slippage test is obtained by applying the appropriate section 1 test to these dummy observations.

The tests can be generalized by replacing the condition that all samples drawn are of the same size by more general conditions. Modifications of this nature can be made in several obvious ways and will not be discussed here. Also in case (a) it is possible to choose r such that $rm > 15$. The restriction $rm \leqq 15$ was imposed so that the tests of Tables 1 and 2 could be used.

GENERALIZED SIGN TEST

A test which has wide application is the well known sign test (see. e.g. [5]). This test is used to compare two kinds of objects (say X and Y) with respect to a specified characteristic. The comparison is accomplished by pairing the two types of objects. In each pair the value of the characteristic obtained for the Y object is subtracted from the value obtained for the X object. This procedure furnishes n dummy observations $x_1-y_1, \cdots, x_n-y_n$, where n is the number of pairs formed. The conditions under which the observations $(x_1, y_1), \cdots, (x_n, y_n)$ where obtained are such that the dummy observations can be assumed to be independent and to have the property.

$$(1)\qquad Pr\,(x_i - y_i > 0) = Pr\,(x_i - y_i < 0) = \tfrac{1}{2}, \qquad (i = 1, \cdots, n),$$

if the null hypothesis of no difference between X and Y with respect to the specified characteristic is true. The sign test is then applied to test this null hypothesis on the basis of the signs (plus or minus) of the n dummy observations.

Let us examine condition (1) from the viewpoint of approximate verification in practice. The main practical situations (i.e., situations approximately satisfied in practice) from which condition (1) can be deduced occur when there is reason to believe that each observation (x_i, y_i) satisfied one or more of the following conditions when the null hypothesis is true:

(a) The joint cumulative distribution $F(x_i, y_i)$ of x_i and y_i is continuous and such that $F(x_i, y_i) = F(y_i, x_i)$; i.e., x_i and y_i receive identical treatment.

(b) x_i and y_i are independent and are from continuous symmetrical populations with the same median value (the two populations are not necessarily the same).

(c) x_i and y_i have the same population median and a normal bivariate distribution.

In all of cases (a)–(c), the dummy observation x_i-y_i is from a continuous symmetrical population with zero median (see [6]). Thus, from the viewpoint of practical verification, assuming that the n dummy observations satisfy condition (1) is almost equivalent to assuming that the dummy observations satisfy conditions (A) with zero median. This suggests that the tests of section 1 be used for those practical situations where the sign test is ordinarily used. The section 1 tests are preferable to the sign test from the viewpoint of suitable significance levels and efficiency for normality.

It would be very convenient if $x-y$ were from a continuous sym-

metrical population with zero median whenever the observation (x, y) satisfies the condition

(d) x and y come from symmetrical populations with the same median value; also the joint cumulative distribution of x and y is continuous.

Then conditions (b) and (c) could be replaced by the single condition (d) and the problem of deciding when $x-y$ is from a continuous symmetrical population with zero median would be considerably simplified. The following counter-example, however, shows that condition (d) is not sufficient to ensure that $x-y$ is from a continuous symmetrical population with zero median:

Let the joint probability density function of x and y be defined by

$$\begin{aligned} f(x, y) &= \frac{1}{4} + \frac{x}{10}(1 - 3y^2) \quad \text{if} \quad -1 \leqq x, y \leqq 1 \\ &= 0 \qquad\qquad\qquad\qquad\quad \text{otherwise.} \end{aligned}$$

Integration shows that the marginal distributions of x and y are both symmetrical with zero median. However $Pr(x-y<0)$ is easily shown to not equal $\frac{1}{2}$.

REFERENCES

[1] John E. Walsh, "Some significance tests for the median which are valid under very general conditions," *Annals of Math. Stat.*, March, 1949.

[2] John E. Walsh, "On the range-midrange test and some tests with bounded significance levels," *Annals of Math. Stat.*, June, 1949.

[3] John E. Walsh, "On the 'information' lost by using a t-test when the population variance is known," *Jour. Amer. Stat. Assoc.*, March, 1949.

[4] Wallace M. Hazel and Warren K. Eglof, "Determination of calcium in magnesite and fused magnesia," *Ind. Eng. Chem. Anal. Ed.*, Vol. 18 (1946), pp. 759–760.

[5] W. J. Dixon and A. M. Mood, "The statistical sign test," *Jour. Amer. Stat. Assoc.*, Vol. 41 (1946), pp. 565–566.

[6] John E. Walsh, "Some significance tests for the median which are valid under very general conditions." Unpublished thesis, Princeton University.

A SAMPLING STUDY OF THE MERITS OF AUTOREGRESSIVE AND REDUCED FORM TRANSFORMATIONS IN REGRESSION ANALYSIS

GUY H. ORCUTT AND DONALD COCHRANE
Department of Applied Economics, Cambridge

This paper is concerned with some aspects of regression analysis when the error terms are autocorrelated and there exists more than one relationship between the variables. In particular, we investigate the merits of autoregressive transformations and the reduced form transformation in dealing with these complications.

An important result is that, unless it is possible to specify something about the intercorrelation of the error terms in a set of relations and to choose approximately the correct autoregressive transformation, a certain amount of scepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observations.

1. INTRODUCTION

THE STATISTICAL estimation of the various parameters which enter into theoretical formulations of economic relationships is one of the main objectives of econometrics and the most common statistical technique used is multivariate regression analysis. The classical method of least squares regression has been shown to give best linear unbiassed estimates of the coefficients when certain well known conditions are fulfilled. If a linear relationship exists between the dependent variable x_{1t} and a set of independent variables $x_2 \cdots x_p$ of the form

$$(1) \qquad x_{1t} = b_0 + \sum_{j=2}^{p} b_{1j}x_{jt} + u_t$$

these conditions are satisfied if among other things[1]

(i) the error term is non-autocorrelated, so that the expected value $E(U_t \cdot U_{t-h}) = 0$ for $h \neq 0$;

(ii) each of the determining variables x_{jt} $(j=2, \cdots, p)$ is independent of the error term U_t, i.e., $E(x_{jt}U_t) = 0$ $(j=2, \cdots, p)$.

[1] For a complete statement of the conditions under which least squares give "best unbiassed" estimates, see F. N. David and J. Neyman, "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs*, Vol. II, London 1938.

The formidable complications which have arisen in estimating the structural parameters of economic relationships have their origin, in so far as they are purely statistical in nature, in the fact that these two conditions are not realistic and have to be relaxed in most applications to economic data. The major complications which have arisen may be classified as follows:

(a) the auto-correlated error complication;
(b) the errors in variables complication;
(c) the simultaneous equations complication.

The first of these complications arises when condition (i), that the errors are independently distributed in time, does not hold while condition (ii), that the determining variables are independent of the error term, cannot be maintained when the other two complications are present.

The simultaneous equations approach. The data used in most formulations of economic relationships are obtained from historical time processes and not from conducted experiments, and as a consequence are the results of the solution of a system of simultaneous relations corresponding to the economic processes involved. In order to obtain accurate estimates of the structural parameters of any single equation it may therefore be necessary to take account of the whole system of simultaneous equations in which it occurs. Consider a simple illustration of this problem. The consumption and price of a commodity enter into both a demand and a supply relation so that if we attempt to find the demand relation by considering only the regression of the quantity consumed on the price of the commodity we are ignoring the fact that price is not an independent variable but will depend on the nature of the supply relation. Haavelmo[2] suggested that in such cases the variables should be considered in a joint normal probability distribution which should be studied to clarify the stochastical relationship which the system of equations implies. Such a method assumes that the errors in the equations are non-autocorrelated and normally distributed and that there are no errors of observation in the variables.

It has been shown that for large samples the parameters estimated by the method of maximum likelihood from this joint probability distribution have certain optimal properties. They are asymptoticallyunbiassed estimates and are also efficient statistics.[3] One method of esti-

[2] T. Haavelmo, "The Probability Approach in Econometrics," Supplement to *Econometrica*, Vol. 12, July 1944.

[3] A more complete discussion can be found in T. Koopmans, "Statistical Methods of Measuring Economic Relationships," *Cowles Commission Discussion Papers* Statistics No. 310 (mimeographed copy of lectures delivered at the University of Chicago 1947) and T. Haavelmo, "Methods of Measuring the Marginal Propensity to Consume," *Journal of the American Statistical Association*, Vol. 42, 1947, pp. 105–122.

mating these parameters is to rewrite the system of equations in the reduced form and solve for each of the endogenous variables in terms of the lagged values of the endogenous variables and the exogenous variables which appear in the system. These solutions will be in terms of linear equations and the method of least squares can then be applied by considering each endogenous variable in turn as the dependent variable. The coefficients in this form possess the properties of best unbiassed estimates for large samples. The structural parameters of the original equations can then be derived from these coefficients but it should be mentioned that it may not always be possible to identify the structural parameters of the original relations from the coefficients of the equations estimated in the reduced form. The problem of identification is a very important one for the method discussed and a careful analysis of the system of equations that is being considered should be made before attempting any statistical application.[4] The estimates obtained from the reduced form method are maximum likelihood solutions for an exactly identified system.

Autocorrelated error terms. In an earlier paper[5] we showed that the error terms involved in many current formulations of economic relationships are highly positively autocorrelated. In doing so, we demonstrated that under these circumstances the application of least squares regression to the original data produced very inefficient estimates of the parameters to be measured and suggested that this efficiency could be recovered by applying an autoregressive transformation to the variables which would make the error term approximately random.

Objects of this paper. In this paper we are concerned with the problem of carrying out regression analysis when the error terms are autocorrelated and there exists more than one relationship between the variables. In particular we investigate the merits of autoregressive transformations and the reduced form transformation in dealing with these complications. It is assumed that there are no errors of observation in the variables.

The problems with which we are dealing are essentially deductive in nature and the ideal solution to them is one reached by purely deductive steps from stated premises. However, since it has not been possible

---

[4] See Koopmans, "Statistical Methods of Measuring Economic Relationships," *op. cit.* A very good description of the practical procedure is contained in M. A. Girshick and T. Haavelmo, "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations," *Econometrica*, Vol. 15, 1947, pp. 79–110. Further references may be found in the various articles to which we have referred.

[5] D. Cochrane and G. H. Orcutt, "Application of Least Squares Regression to Relationships containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 32–61.

as yet to obtain such a solution, and considering the problems of some importance, we have resorted to the method of sampling experiments. That is, we embody our assumptions in experimental models, use these models to generate sets of time series and investigate empirically the results of various estimating procedures on the series generated. Up to the present only large sample properties of the parameters derived by the simultaneous estimation of structural relations have been demonstrated. Since economic data rarely comprise series of more than 20 years, these large sample properties would seem to require more careful investigation and the sampling experiment approach provides a convenient and legitimate method of making such an investigation. In addition it might be mentioned that the use of such methods might also provide the answers to many problems which have proved intractable to mathematical statistics and the improvements in calculating equipment are very welcome for these purposes.

2. CONSTRUCTION OF THE EXPERIMENTAL MODELS

In order to reduce the computational burden as much as possible we worked with the simplest types of systems which seemed at all reasonable from the standpoint of applying any conclusions to economic studies. Two models were adopted and are explained as follows. The original series of Model I were generated by a recursive system of equations

$$x_t = a_0 + a_1 y_t + (u_{1t} + u_{2t}) \tag{2}$$

$$y_t = b_0 + b_1 x_{t-1} + u_{3t} \tag{3}$$

where x_t and y_t are the series to be considered and u_{it} $(i=1, 2, 3)$ are the error terms involved in the two relations. These error terms were generated by the autoregressive equations

$$u_{it} = u_{i,t-1} + \epsilon_{it} \qquad (i = 1, 2, 3) \tag{4}$$

where the ϵ_{it} denote series of random disturbances. The values of the parameters in (2) and (3) were chosen to be

$$\begin{aligned} a_0 &= b_0 = 0 \\ a_1 &= 1.0 \\ b_1 &= 0.4 \end{aligned} \tag{5}$$

The ϵ_{it} $(i=1, 2, 3)$ are independently distributed single digit random numbers. They were extracted from *Tables of Random Sampling*

Numbers,[6] ignoring zeros so that they ranged from 1 to 9. Subtracting the number 5 from each we obtained three random series ϵ_{it}, possessing rectangular distributions with ranges of $+4$ to -4 and an expected value of zero. The three series u_{it} were then generated by applying equations (4) with initial values of zero, so that we had three independent series of first summations of random elements each comprising over 500 terms. Taking x_0 as zero and making use of the properties of the system given by (2) and (3), we then generated long series of x_t and y_t from the three error series u_{it}. The first five items of the series of x_t and y_t were discarded and the remaining long series were each divided into 20 segments of 21 items with 5 items omitted between segments. One of these items was later used for prediction. By this procedure we obtained a sample of 20 pairs of series generated by the same underlying autoregressive structure but involving different samples of random disturbances.

For Model II the original series were generated by the same process as just described for Model I except that instead of being independent the error terms were now highly intercorrelated. This result was achieved by using the same series of u_{1t} and u_{2t} as before but replacing the series u_{3t} by u_{2t} so that the true correlation between the error terms was 0.71. The recursive system therefore became

$$x_t' = a_0 + a_1 y_t' + (u_{1t} + u_{2t}) \tag{6}$$

$$y_t' = b_0 + b_1 x_{t-1}' + u_{2t} \tag{7}$$

where the constants remained the same as given in (5). The same procedure was used to obtain the individual sets of x_t and y_t as explained for Model I.

Choice of parameters. Our choice of the autoregressive properties of the error series u_{it} was based upon the evidence presented in our previous paper[7] and the reasonableness of assuming that error terms are first summations of random elements has been further supported from the results obtained by Stone for a number of demand studies in the United Kingdom.[8] Our choice of the product a_1b_1 was made so that x_t and y_t would have approximately the same autoregressive

---

[6] M. G. Kendall and B. Babington-Smith, "Tables of Random Sampling Numbers," *Tracts for Computors No. 24*, Cambridge University Press 1939.

[7] D. Cochrane and G. Orcutt, *op. cit.*

[8] Richard Stone, "The Analysis of Market Demand: An Outline of Methods and Results" read before a meeting of the European section of the Econometric Society at The Hague, September 1948, and to be published in *The Review of the International Statistical Institute.*

structures as claimed by Orcutt[9] for the series used in Tinbergen's[10] model of the economic system of the United States. For instance in the case of independent error terms we can see from (2) and (3)that the autoregressive structures of the two series are

$$x_t = x_{t-1} + 0.4(x_{t-1} - x_{t-2}) + \eta_{1t} \tag{8}$$

$$y_t = y_{t-1} + 0.4(y_{t-1} - y_{t-2}) + \eta_{2t} \tag{9}$$

where η_{1t} and η_{2t} are random disturbances defined in terms of $\epsilon_{it}(i = 1, 2, 3)$.

Having made the decision as to the product a_1b_1, only one more significant decision remains to be made about the general structure of the model. This is the correlation between either pair of variables x_t and y_t or y_t and x_{t-1}. For n approaching infinity this may be more clearly seen as follows. We may express our model in the form

$$x_t = ay_t + v_{1t} \tag{10}$$

$$y_t = bx_{t-1} + v_{2t} \tag{11}$$

where x_t and y_t are in terms of deviations from their means and v_{1t}, v_{2t} are random error series. Expressing x_t and y_t in autoregressive forms, we can derive the following relations:

$$\frac{E(x_t^2)}{E(v_{1t}^2)} = \frac{1}{1 - \rho^2}\left(1 + a^2 \frac{E(v_{2t}^2)}{E(v_{1t}^2)}\right) = R_1 \tag{12}$$

$$\frac{E(y_t^2)}{E(v_{2t}^2)} = \frac{1}{1 - \rho^2}\left(\frac{\rho^2}{a^2}\, \frac{E(v_{1t}^2)}{E(v_{2t}^2)} + 1\right) = R_2 \tag{13}$$

where

$$ab = \rho, \quad R_1 = \frac{1}{1 - r_{x_t y_t}{}^2} \quad \text{and} \quad R_2 = \frac{}{1 - r_{y_t x_{t-1}}{}^2}$$

It can be readily seen that, having decided the correlation between say x_t and y_t, we have R_1 and since the term

$$a^2 \frac{E(v_{2t}^2)}{E(v_{1t}^2)}$$

[9] G. H. Orcutt, "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States 1919–32," *Journal of the Royal Statistical Society*, Vol. X, Series B, 1948, pp. 1–53.

[10] J. Tinbergen, "Statistical Testing of Business-Cycle Theories Vol. II; Business Cycles in the United States of America 1919–32," League of Nations, Geneva, 1939.

appears in both (12) and (13), then R_2 is a function of ρ and R_1 which are both known and is automatically determined. All we have left to decide is the weights to be assigned to the coefficient a and the relative variances of the error terms. We made the relative variances

$$E(v_{2t}^2)/E(v_{1t}^2) = \tfrac{1}{2},$$

so that for $r_{x_t y_t}^2 = 0.44$ we have $a = a_1 = 1$. The resulting form provides an intermediate and reasonable model of a simplified economic system. On the basis of the calculations needed for this study, it is possible to work out the extreme cases in which either of the error terms has zero variance. This is done in the next section.

3. CALCULATIONS INVOLVED AND SOME SPECIAL CASES

The results of the calculations carried out are contained in Tables 1 to 5. The equations referred to in these tables are:

Model I.

$$\text{(14)}\qquad \begin{aligned} &\text{I.} & x_t &= a_0 + a_1 y_t + (u_{1t} + u_{2t}) \\ &\text{II.} & y_t &= b_0 + b_1 x_{t-1} + u_{3t} \\ &\text{III.} & x_t &= \rho_0 + \rho_1 x_{t-1} + (u_{1t} + u_{2t} + a_1 u_{3t}). \end{aligned}$$

Model II.

$$\text{(15)}\qquad \begin{aligned} &\text{IV.} & x_t' &= a_0 + a_1 y_t' + (u_{1t} + u_{2t}) \\ &\text{V.} & y_t' &= b_0 + b_1 x_{t-1} + u_{2t} \\ &\text{VI.} & x_t' &= \rho_0 + \rho_1 x_{t-1}' + (u_{1t} + (1 + a_1) u_{2t}). \end{aligned}$$

Both these systems are exactly identified. Equations I and IV were calculated by the direct application of least squares regression. Equations II and V are already in the reduced form so that the use of least squares is the appropriate procedure. The reduced forms of equations I and IV are equations III and VI respectively, so that the reduced form estimates of a_1 and a_0 are given by

$$\text{(16)}\qquad a_1' = \rho_1 / b_1$$

$$\text{(17)}\qquad a_0' = \rho_0 - b_0 a_1' = \bar{x}_t - a_1' \bar{y}_t$$

where $\bar{x}_t$ and $\bar{y}_t$ are the means of the two series x_t and y_t. The calculations relating to these estimates are given by equations IA and IVA in Model I and Model II respectively. For each of the equations we have made first and second difference transformations and the regression parameters have been estimated in the three forms. The original

relations and the autoregressive transformations are denoted by the letters O, F.D. and S.D. in the tables.

Special cases. At the end of section 2 we pointed out that it is possible to derive the cases where the variance of either one of the error terms is equal to zero. In both these cases the least squares estimates and the reduced form estimates lead to identical results. Rewriting the simple system of (10) and (11)

$$x_t = ay_t + v_{1t} \tag{18}$$

$$y_t = bx_{t-1} + v_{2t} \tag{19}$$

where v_{1t}, v_{2t} are random elements so that the reduced form of x_t is

$$x_t = \rho x_{t-1} + v_{1t} + av_{2t} \qquad (\rho = ab) \tag{20}$$

we can say that when (19) is an exact relation then the single equation least squares estimates and the reduced form estimates of a are the same and proportional to the estimate of the autoregressive coefficient obtained in (20). When (18) is an exact relation then the single equation least squares and the reduced form estimates of a are both exact. These equalities may be more clearly seen as follows:

First assumption $E(v_{2t}^2)=0$
so that our system becomes

$$x_t = ay_t + v_{1t} \tag{21}$$

$$y_t = bx_{t-1} \tag{22}$$

$$x_t = \rho x_{t-1} + v_{1t} \tag{23}$$

and the least squares estimate of b is exact.

The single equation least squares estimate of a is given by

$$\hat{a} = \frac{\sum x_t y_t}{\sum y_t^2} = \frac{1}{b}\,\frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2} = \frac{\hat{\rho}}{b} \tag{24}$$

where $\hat{\rho}$ is the least squares estimate of ρ. The reduced form estimate of a is therefore

$$\bar{a} = \frac{\hat{\rho}}{b} = \hat{a}. \tag{25}$$

Second assumption $E(v_{1t}^2)=0$
so that our system is now

$$x_t = ay_t \tag{26}$$

$$y_t = bx_{t-1} + v_{2t} \tag{27}$$

$$x_t = \rho x_{t-1} + av_{2t} \tag{28}$$

and the single equation least squares estimate of a is exact. Now (27) and (28) are identical except for a scalar multiplier a, therefore the reduced form estimates of a are $\hat{p}/\hat{b}$ and will also be exact.

4. GENERAL RESULTS

In the ensuing discussion we shall be content to point out the general and more important features of the calculations and if the reader desires further information it may be obtained from the tables which are presented in detail.

Structural parameters. It has been proved by Mann and Wald[11] that for large samples a linear stochastic difference equation may be treated as a classical regression problem in which the lagged values of the series appear as independent variables. However, the adequacy of the ordinary least squares regression has not been demonstrated for small samples, particularly of the size usually considered by economists, and in fact Koopmans[12] has mentioned that for a sample of 3 items are bias will be present in the least squares estimates of the parameters of a single lag autoregressive equation. Orcutt[13] has pointed out that this bias is probably due partly to the necessity of using the sample means of the time series instead of the true means and partly to the skewness of the distribution of sample estimates even when the true means are used. His empirical evidence shows that this bias is very substantial for series having only a weak central tendency. If we look at equations III and VI in Table 1 we find, for the first difference transformation, examples of single lag autoregressive equations. In these cases the means of the estimated regression coefficients are given, and it can be seen that the biasses are 2.3 and 2.8 times the standard error of the means of twenty estimates. This indicates that even for low values of the autoregressive coefficient the bias is still rather large. When we estimate the coefficients assuming a true mean of zero we find from equations III and VI in Table 2 that the bias is considerably reduced but is still far from negligible.

The question naturally arises as to whether a similar sort of small sample bias is to be expected in single equation least squares and reduced form estimates of the parameters of a system of recursive equations. In the previous section we showed that when either one of the equations was exact the estimates of both methods were the same. They were exact in one case and possessed the same bias and variance as the

[11] H. B. Mann and H. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, Vol. 11, 1943, pp. 173–220.

[12] T. Koopmans, "Serial Correlation and Quadratic Forms in Normal Variables," *Annals of Mathematical Statistics*, Vol. 13, 1942, pp. 14–33.

[13] G. H. Orcutt, *op. cit.*

Equation	Autoregressive transformation		Constant term (True value zero)			Regression Coefficient						Correlation coefficient			Estimated mean variance of error
	Single equation	Reduced form	Mean	Standard error of mean	Variance	True value	Mean	Standard error of mean	Variance: Using mean	Variance: Using true value	Mean of estimated variance	Mean	Standard error of mean	Variance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Model I															
I	O		−25.75	13.62	3712.	1.0	1.27	0.11	0.223	0.283	0.038	0.84	0.05	0.057	20.6
	F.D.		− 0.18	0.16	0.54	1.0	0.97	0.05	0.053	0.052	0.082	0.62	0.03	0.014	14.1
	S.D.		− 0.04	0.08	0.11	1.0	0.61	0.07	0.094	0.239	0.129	0.40	0.03	0.021	28.2
IA		O	35.11	116.91	298030.	1.0	2.10	0.48	4.502	5.494	—	—	—	—	83.4
		F.D.	− 0.63	1.05	30.11	1.0	0.62	0.15	0.432	0.551	—	—	—	—	18.1
		S.D.	0.18	0.69	10.15	1.0	−1.82	0.66	3.605	16.104	—	—	—	—	198.3
II	O		−10.57	9.55	1824.	0.4	0.59	0.05	0.058	0.091	0.007	0.82	0.05	0.045	11.4
	F.D.		− 0.25	0.15	0.43	0.4	0.36	0.03	0.014	0.015	0.017	0.54	0.03	0.018	6.8
	S.D.		− 0.07	0.03	0.02	0.4	0.20	0.03	0.021	0.061	0.023	0.29	0.04	0.037	12.3
III	O		−18.36	4.71	444.	0.4	0.90	0.02	0.008	0.255	0.013	0.88	0.02	0.006	22.8
	F.D.		− 0.57	0.24	1.19	0.4	0.26	0.06	0.062	0.079	0.050	0.25	0.05	0.057	21.4
	S.D.		0.01	0.09	0.15	0.4	−0.27	0.05	0.050	0.490	0.053	−0.27	0.05	0.049	30.3
Model II															
IV	O		−14.28	7.77	1206.	1.0	1.48	0.06	0.079	0.301	0.016	0.94	0.01	0.004	18.3
	F.D.		− 0.07	0.10	0.183	1.0	1.51	0.03	0.020	0.276	0.054	0.83	0.02	0.005	11.2
	S.D.		− 0.03	0.07	0.097	1.0	1.53	0.03	0.072	0.345	0.140	0.70	0.02	0.006	28.2
IVA		O	− 8.23	9.59	1839.	1.0	1.52	0.08	0.114	0.378	—	—	—	—	14.4
		F.D.	− 0.54	0.26	1.334	1.0	0.52	0.18	0.668	0.863	—	—	—	—	24.2
		S.D.	3.03	2.04	83.038	1.0	5.23	7.85	1226.5	1883.1	—	—	—	—	17470.1
V	O		− 8.07	5.94	760.	0.4	0.59	0.03	0.018	0.055	0.004	0.89	0.02	0.007	10.0
	F.D.		− 0.13	0.12	0.279	0.4	0.33	0.02	0.011	0.014	0.011	0.59	0.03	0.020	7.4
	S.D.		− 0.06	0.05	0.042	0.4	0.09	0.02	0.006	0.099	0.012	0.19	0.03	0.020	11.7
IV	O		−28.13	5.34	570.	0.4	0.86	0.02	0.009	0.224	0.014	0.86	0.02	0.005	36.2
	F.D.		− 0.33	0.23	1.082	0.4	0.23	0.06	0.062	0.086	0.050	0.23	0.06	0.061	34.9
	S.D.		− 0.15	0.11	0.249	0.4	−0.34	0.04	0.036	0.583	0.051	−0.34	0.04	0.034	47.7

coefficient of a single lag autoregressive equation of the type just considered in the other. It is therefore of interest to examine first our Model I which corresponds to an intermediate case from the two extremes and second our Model II which adds the complication due to intercorrelated error terms and accordingly a further bias to the single equation estimates by the direct correlation between the independent variable and the error series.

First look at the single equation estimates of the regression coefficients given in Table 1, by the sets of equations I and IV. The estimates based on the original series are badly biassed and have large variances in both methods. As expected the bias is greater in Model II where the error terms are intercorrelated. The mean of the variances of the regression coefficients estimated from each equation separately (see column 11) does not reflect the true position and is only a fraction of what it should be. When we make a first difference transformation the estimates of the regression coefficient in Model I possess very little bias while the mean of the estimated variances of the regression coefficients also appears to be reasonable. However, in Model II there is still a large bias due to the correlation of the independent variable and the error term of equation IV.

Turning to the reduced form estimates given by equations IA and IVA we see that they are badly biassed for both the original series and first differences of both models. In the case of the original series the

TABLE 2
REGRESSION PARAMETERS CALCULATED BY ASSUMING TRUE MEAN OF ZERO
(First Difference Transformation)

Equation	Regression Coefficient					Correlation Coefficient		
	True Value	Mean	Standard error of mean	Variance: Using mean	Variance: Using true value	Mean	Standard error of mean	Variance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model I								
I	1.0	1.00	0.05	0.041	0.039	0.66	0.03	0.014
IA	1.0	0.79	0.12	0.290	0.319	—	—	—
II	0.4	0.39	0.03	0.015	0.014	0.58	0.03	0.017
III	0.4	0.33	0.05	0.059	0.061	0.33	0.05	0.056
Model II								
IV	1.0	1.50	0.03	0.021	0.271	0.84	0.01	0.004
IVA	1.0	0.66	0.13	0.352	0.446	—	—	—
V	0.4	0.35	0.02	0.009	0.011	0.61	0.03	0.017
IV	0.4	0.27	0.05	0.056	0.070	0.27	0.05	0.055

biasses are in the same direction as the single equation biasses but are much larger, in the first difference transformation the bias is downwards in both models and is due to the short series bias previously considered.

So far we have considered the coefficients obtained when we estimated the means in each transformation. It is therefore of interest to see the bias in the estimates of the first difference transformation when we make use of the fact that the true mean of the series is zero. This is equivalent to assuming that there is no trend in the original relationships. The results for the first difference transformation of both models are given in Table 2. In the case of Model I the estimates obtained by the single equation least squares regression are not biassed but the same bias as previously obtained is present in the case of equation IV for Model II. The reduced form estimates are still biassed in both models, although they show an improvement over the estimates obtained when the means are estimated.

Our calculations may be used to see whether the means of the reduced form estimates are significantly biassed from the true values in the first difference transformations. When we calculate the coefficient using estimated means we find from Table 1 that the values 0.62 from equation IA and 0.52 from equation IVA are both significantly different from the true value of the regression coefficient at the 5 per cent. level,[14] using the standard error of the mean calculated from the variance around the estimated mean. In fact they are significantly different from the true value of the coefficient at the 2 per cent level. When we assume that the true means of the series are zero only the mean of the reduced form estimates for Model II is significantly biassed. From Table 2 it can be seen that the value of 0.79 for equation IA is not significant at the 5 per cent. level but the value of 0.66 for equation IVA is significant at the 2 per cent level.

Consider now the efficiency of the least squares and reduced form estimates. In Table 1 we find that the variances of the reduced form estimates compare very unfavourably in all cases with the variances of the single equation estimates. This may be illustrated by the ratios of the reduced form variances to the single equation variances. Even when we include the effect of the bias on the estimates by calculating the variance around the true value of the regression coefficient, the ratios still remain very high for Model I and although they fall slightly in Model II they are still greater than unity. For the cases

[14] Using the t tables for 20 degree of freedom. We are using both tails of the distribution but a case can be made for using only one tail. This would halve the levels of significance.

	Ratios of variances regression estimates	
	Using estimated mean	Using true value
Model I		
Original series	20:1	19:1
First differences	8:1	11:1
Model II		
Original series	$1\frac{1}{2}$·1	$1\frac{1}{2}$:1
First differences	30:1	3:1

where we assume a knowledge of the true mean, Table 2 shows that the ratio of the variances of the reduced form estimates to the variances of the single equation estimates around the estimated means and the true value are 7:1 and 8:1 respectively for Model I and 17:1 and $1\frac{1}{2}$:1 for Model II. These are still very high ratios and are surprising results.

In Model I, where the error terms are random and independent, the single equation estimates would be the maximum likelihood solutions for large samples and normally distributed error terms[15] and we therefore expected the single equation approach to produce better estimates than those obtained by the reduced form method. However, where the error terms are random and intercorrelated the reduced form estimates are the maximum likelihood solutions for large samples if the correlation between the unlagged error terms is unknown. This is the case of Model II and we should therefore have expected the results of Model I to be reversed, but the discussion of the last few paragraphs has shown that these expectations have been far from realised.

Autocorrelation of residuals. The autocorrelations of the true error terms and the estimated residuals are presented in Table 3. The measure of autocorrelation used is the ratio of the mean square successive difference to the variance. This statistic is usually denoted by δ^2/s^2 and for a random series it is symmetrically distributed around a mean of $2n/n-1$ where n is the number of items in the series.[16] In a previous paper[17] we showed that highly positively autocorrelated error terms become strongly biassed towards randomness as the number of para-

---

[15] R. Bentzel and H. Wold, "On Statistical Demand Analysis from the viewpoint of Simultaneous Equations," *Skandinavisk Aktuarietidskrift*, Vol. 29, 1946, pp. 95–114.

[16] The probability distribution of this statistic has been tabulated; see B. S. Hart and J. von Neumann, "Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Varian," *Annals of Mathematical Statistics*, vol. 13, pp. 207–214.

[17] D. Cochrane and G. Orcutt, *op. cit.*

meters in the estimation relationship increases. This result is again illustrated in Table 3.

Further it should be noted that not only are the residuals biassed when the single equation least squares method of estimation is used, but this bias appears to be approximately the same as can be seen from (column 7) Table 3, when the reduced form method of estimation is used.

TABLE 3
AUTOCORRELATION OF RESIDUALS

Equation	Autocorrelation transformation		Values of δ^2/s^2						
			Actual error series				Estimated residuals		
	Single equation (1)	Reduced form (2)	Values for infinite series (4)	Mean (5)	Standard error of mean (5)	Variance (6)	Mean (7)	Standard error of mean (8)	Variance (9)
Model I									
I	O		0.0	0.61	0.09	0.14	1.00	0.10	0.22
	F.D.		2.0	2.16	0.10	0.20	2.21	0.10	0.21
	S.D.		3.0	—	—	—	—	—	—
IA		O	0.0	0.61	0.09	0.14	1.07	0.10	0.19
		F.D.	2.0	2.16	0.10	0.20	1.93	0.09	0.18
		S.D.	3.0	—	—	—	—	—	—
II	O		0.0	0.43	0.06	0.08	0.97	0.11	0.22
	F.D.		2.0	2.05	0.09	0.17	2.00	0.08	0.14
	S.D.		3.0	—	—	—	—	—	—
III	O		0.0	—	—	—	1.49	0.10	0.21
	F.D.		2.0	—	—	—	1.87	0.05	0.04
	S.D.		3.0	—	—	—	—	—	—
Model II									
IV	O		0.0	0.61	0.09	0.14	1.10	0.13	0.34
	F.D.		2.0	2.16	0.10	0.20	2.50	0.10	0.20
	S.D.		3.0	—	—	—	—	—	—
IVA		O	0.0	0.61	0.09	0.14	1.04	0.11	0.24
		F.D.	2.0	2.16	0.10	0.20	2.07	0.11	0.24
		S.D.	3.0	—	—	—	—	—	—
V	O		0.0	0.51	0.06	0.08	1.17	0.10	0.18
	F.D.		2.0	2.17	0.09	0.17	2.01	0.08	0.11
	S.D.		3.0	—	—	—	—	—	—
IV	O		0.0	—	—	—	1.47	0.11	0.22
	F.D.		2.0	—	—	—	1.92	0.06	0.06
	S.D.		3.0	—	—	—	—	—	—

There is also an additional bias in the residuals caused by the application of biassed estimates of the regression coefficients. This may be seen in the residuals of the first difference transformation in equa-

tion IV, Table 3 (column 7). It can be easily shown that the application of biassed coefficients in this equation will produce negatively autocorrelated error terms and this is illustrated in the result obtained. A final interesting feature of the autocorrelation of the residuals is to be seen in equations III and VI. In the first difference transformation we are estimating the coefficient of a single lag autoregressive equation with random disturbance but it is noticeable that because of the downward bias in the estimate of the coefficient the residuals are not completely randomized.

Prediction. In each equation we forecasted the dependent variable

TABLE 4

ERRORS OF FORECAST

Equation	Autoregressive transformation		Actual errors of forecast			Estimated mean variance of	
	Single equation (1)	Reduced form (2)	Mean (3)	Standard errors of mean (4)	Variance (5)	Error (6)	Individual forecast (7)
Model I							
I	O		2.45	1.09	23.8	20.6	24.0
	F.D.		−0.03	0.65	8.6	14.1	16.2
	S.D.		−1.52	1.10	24.3	28.2	31.3
IA		O	6.19	4.57	417.1	83.4	—
		F.D.	−1.03	0.76	11.5	18.1	—
		S.D.	−5.86	3.43	235.7	198.3	—
II	O		−1.27	1.04	21.6	11.4	13.8
	F.D.		−1.06	0.67	9.0	6.8	7.6
	S.D.		−1.12	0.68	9.2	12.3	13.9
III	O		0.63	1.35	36.6	22.8	26.7
	F.D.		−1.18	1.04	21.9	21.4	23.8
	S.D.		−2.30	1.18	27.8	30.3	34.0
Model II							
IV	O		1.92	1.00	25.6	13.3	15.4
	F.D.		−0.66	0.59	7.0	11.2	12.4
	S.D.		−1.22	1.19	28.2	28.2	31.0
IVA		O	1.65	1.14	26.0	14.4	—
		F.D.	0.38	0.91	16.4	24.2	—
		S.D.	40.56	34.39	23654.	17470.	—
V	O		0.38	0.74	10.9	10.0	11.6
	F.D.		0.02	0.53	5.6	7.4	8.3
	S.D.		−0.77	0.68	9.2	11.7	13.1
IV	O		2.10	1.07	23.0	36.2	41.4
	F.D.		0.31	1.03	21.0	34.9	38.9
	S.D.		−1.96	1.24	30.7	47.7	53.9

for the next period from a knowledge of the independent variable and using the regression coefficients calculated from the series up to that period. The forecasts were then compared with the actual value of the dependent variable obtained from our constructed series as explained in section 2, and the variances of error are shown in Table 4. The forecasts based directly upon the single equation estimates of the parameters have a smaller variance of error than those based on the direct use of the reduced form estimates of the parameters in the set of structural equations IA and IVA. This is true for both models and each autoregressive transformation. In both models the smallest variance of error of forecast is given when using single equation estimates in the first difference transformation.

We also calculated the variance of forecast for the two main equations of both models in first difference transformation using the knowledge that the true means of the series were zero. Table 5 shows that these results do not change the general impression and the single equation least squares estimates still provide the smallest variances of the errors of forecast.

TABLE 5

ERRORS OF FORECAST ASSUMING TRUE MEAN OF ZERO

(First Difference Transformation)

Equation	Actual errors of forecast		
	Mean	Standard error of mean	Variance
Model I			
I	−0.10	0.58	6.66
IA	0.36	0 66	8.64
Model II			
IV	0.41	0.54	5 82
IVA	−0 69	0.90	16 15

5. CONCLUSIONS

Care must be taken when drawing general conclusions from sampling experiments of the type considered in this paper but the results appear to us to be of a striking and significant nature. They indicate that unless it is possible to specify with some degree of accuracy the intercorrelation between the error terms of a set of relations and unless it is possible to choose approximately the correct autoregressive transformation so as to randomize the error terms, then a certain amount of scepticism is justified concerning the possibility of estimating structural parameters from aggregative time series of only twenty observa-

tions when generated by systems analogous to those examined in this paper. This scepticism will be considerably increased if it is also attempted to make a choice of variables and time lags from the same data

If the error terms are independent or nearly independent between equations, the results obtained justify the use of single equation least squares estimation, in which case the main problem lies in making the correct autoregressive transformation. This would seem to be the situation as regards many demand relationships, particularly those which are in terms of current variables and those relating to agricultural products.

For short run prediction we would appear to be in a somewhat more favourable position since the estimated variance of errors of individual forecasts obtained from single equation regression analysis seem to be in line with the actual errors and as small as could be expected even in the absence of a simultaneous equations complication. However, it must be remembered that such predictions assume that the same system will continue.

CONTROL OF A GENERAL CENSUS BY MEANS OF AN AREA SAMPLING METHOD

Gabriel Chevry
Institut National de la Statistique et des Études Économiques, Paris

The area sampling method was applied to industrial and commercial enterprises in urban areas of France to determine the adequacy of the general census in enumerating them. The investigation showed an under-reporting by the census of enterprises of all sizes.

A general census of the population was taken in France on March 10, 1946. Like the previous ones, this census involved in effect several simultaneous inquiries about people, households, dwellings, industrial and commercial enterprises, and agricultural establishments; different questionnaires were used for each of these categories of statistical units.

Questionnaire No. 5 applied to the industrial and commercial enterprises answering the following definition: an enterprise consists of a group of two or more persons working together permanently, in a permanent place, under the management of one or more representatives of the same trade name. According to this definition, a person working alone did not constitute an enterprise. Two partners, or a husband and wife working together, without assistants, constituted an enterprise employing no salaried staff. Each of the various subsidiaries of the same firm constituted a separate enterprise, even if they were located in the same community.

A first reckoning of the questionnaires No. 5 which had been filled through the census procedure revealed some anomalies, which indicated that more than a negligible number of enterprises had not filled their answers to the census questionnaires. In order to have a more precise impression and to be able to appraise quantitatively the accuracy of the census of industrial and commercial enterprises, and management of the Direction des Enquetes economiques of the Institut national de la Statistique decided in October 1946 to resort to its regional offices for an *a posteriori* control of the census process by means of an area sampling method.

PRINCIPLES OF THE METHOD

To reduce the cost of the process, the control was limited to the

urban areas in which the Institut national de la Statistique has regional offices. It could be presumed that the census of March 10, 1946, had been more accurately taken in the rural areas than in the large cities.

The control consisted simply in selecting at random in each of these cities, a certain number of areas in which the sampling was to take place, in listing the enterprises located in the selected areas, and in sending investigators to visit all these areas to make an inventory of all the enterprises located therein and to obtain completed questionnaires No. 5 from those enterprises which had not submitted their answers to the census.

For each city the number of areas selected was considered adequate when, according to the census data, the number of enterprises located in these areas was at least equal to 2.5 per cent of the total number of enterprises located in the whole urban center.

TECHNIQUE OF SAMPLING

The first step consisted in setting the limits of the urban centers where the sampling was to take place. It was required that the sampling not be limited to the city in which the regional office of the Institut national was located, but should extend to the whole urban center in which the city was included, this center consisting of all the small communities or bounding parts of these communities in which the inhabited lots of these divisions were adjoining or joined together by parks, gardens, orchards, yards, workshops or other similar enclosures, even if these houses or enclosures were separated from each other by a ditch, a river or a public garden.

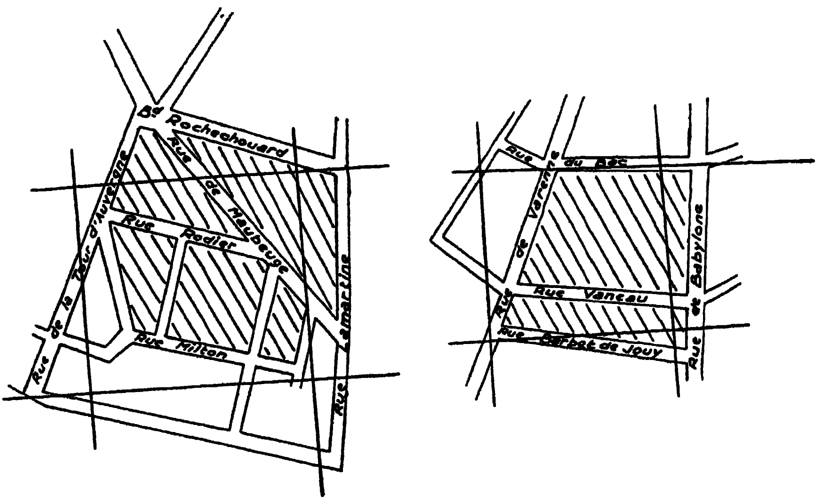
A large scale map of the urban center thus delineated was covered with a regular orthogonal grid forming squares of 250 meters per side for the smaller centers and 500 meters per side for the very large ones. The squares covering areas containing construction were numbered and a certain number of squares were taken at random by means of Tippett tables.

As it was impossible to locate on the terrain the limits of the selected grids, instructions were given to substitute for each of these grids, polygons of approximately equal areas but bounded by actual streets. The chart below shows two examples of this substitution, which was indispensable in order to: a) facilitate the listing of the enterprises located in each of the grids which had filled the questionnaire; and b) point out unambiguously the buildings which were to be visited by the investigators.

All the buildings located in the polygons thus selected and delineated

were visited by the investigators. Each of the investigators was provided with a list of the enterprises which, according to the census, were located in the group of buildings assigned to him. He checked on this list the enterprises which had completed answers to Questionnaire No. 5 for the census authorities, and those which had not been polled by the census.

As this procedure took place in November 1946, several months after the general census, special instructions were given to the investigators,



SUBSTITUTION OF POLYGONS FOR SQUARES. AREAS INCLUDED IN POLYGONS ARE SHOWN AS SHADED REGIONS

so that they could distinguish, among all of the enterprises existing in November 1946 and not having replied to the census questionnaire, those which had been in existence in March 1946 and thus should have been included in the census list, and those which had been created or reopened between March and November. Conversely, it was hardly possible to discover in November the number of enterprises which had been closed since March. This fact seems of no consequence, however, as bankruptcies were very scarce in this period after the war.

RESULTS OF THE SAMPLING

Table 1 shows the gross results of the sampling for each of the selected cities and for the sample as a whole.

In the 18 urban centers where the sampling control was used,

TABLE 1

GROSS RESULTS OF THE SAMPLE

Urban Centers	Total number of enterprises included in the census of March 1946 in these centers	Number of enterprises existing in the sample: Included in the census of March 1946 — Number	As a percentage of the total number of enterprises in the census	Not included in the census of March 1946 — Number	As a percentage of the number of enterprises in the sample
(1)	(2)	(3)	(4) $=\frac{(3)}{(2)}$	(5)	(6) $=\frac{(5)}{(3)+(5)}$
			%		%
Bordeaux	10,556	365	3.4	97	21.0
Clermont	3,185	127	4.0	52	29.1
Dijon	2,366	60	2.5	29	32.6
Lille	7,549	228	3.0	188	45.2
Limoges	2,870	105	3.6	34	24.5
Lyon	14,230	405	2.8	369	47.6
Marseille	14,355	531	3.7	182	25.5
Montpellier	3,281	161	4.9	61	27.5
Nancy	3,272	88	2.7	27	23.5
Nantes	5,005	195	3.9	19	8.9
Orleans	2,748	109	3.9	17	13.5
Paris	110,000	3,115	2.8	529	14.5
Poitiers	905	53	5.9	38	41.8
Reims	3,233	120	3.7	9	7.0
Rennes	2,835	89	3.1	23	20.5
Rouen	5,351	160	3.0	87	35.2
Strasbourg	5,958	128	2.2	60	31.9
Toulouse	4,580	115	2.5	82	41.6
Total	201,359	6,154	3.1	1,903	23.6

201,359 enterprises had been included in the census of March 1946. The selected sample included 6,154 enterprises, i.e. 3.1 per cent of the total.

The investigation made on the spot has shown that 8,057 enterprises existed in this sample. This means that 1,903 enterprises, or an amount equal to 30.9 per cent of the enterprises included in the census and 23.6 per cent of the existing enterprises, had been overlooked in the census of March 1946.

Sampling theory permits one to determine the accuracy of this gross result. It does not seem possible to use unreservedly the formula which, in the case of a cluster sample gives the variance of an estimate of probability, p; (p, in this case, would be the probability of an enterprise being overlooked in the census). This formula supposes in effect

that all the clusters are of the same size, i.e. include the same number of enterprises. It is obvious that, with the method applied in the sample, this condition has not been fullfilled.

To measure the accuracy of the results under these circumstances, it seems more advisable to apply the approximate formula provided by Goldberg.[1] This formula deals with the errors in estimating a ratio x/y, when the ratio is obtained as the quotient of estimates of its numerator and denominator, both computed from at least n elements. This estimate implies a bias whose average relative value is

$$\frac{C_x^2 - \rho C_x C_y}{n}$$

and the relative standard error of sampling is

$$\sqrt{\frac{C_x^2 + C_y^2 - 2\rho C_x C_y}{n}}$$

where

C_x is the coefficient of variation of x, namely $\sigma_x/\bar{x}$,
$\bar{x}$ is the arithmetical mean of the values of x in the universe C_y,
C_y is the coefficient of variation of y namely $\sigma_y/\bar{y}$, and
ρ is the coefficient of correlation between x and y.

In the particular case which we are considering, n is the number of polygons, bounded by streets, which were taken at random; the values of y are the number of the enterprises existing in each of these polygons, and the values of x are the number of enterprises overlooked by the census in the same polygons.

The coefficient of correlation is rather high: 0.91. The average relative value of the bias is 2/1000, i.e. practically negligible. The relative value of the standard error of the sample is 7.6 per cent.[2]

It can therefore be estimated that in 95 cases out of 100 the percentage of enterprises overlooked in the census lies within the limits of 23.6(1−2×0.076) or 20 per cent, and 23.6(1+2×0.076) or 27.2 per cent. Thus we can state that, in the census of March 10, 1946, 20 to 27 per cent of the industrial and commercial enterprises were overlooked.

It has already been pointed out that *a priori* it does not seem sound

[1] Cf: "Méthodes statistiques modernes des Administrations fédérales aux États-Unis" by P. Thionet (Hermann & Cie, Paris, 1946) p. 58, and "Sampling theory when the sampling units are of unequal sizes" by W. G. Cochran (*Journal of the American Statistical Association*, June 1942).

[2] Actually, these calculations could be made for only 14 out of the 18 cities in which the sampling took place. In the other 4 cities, we could not use the results relative to each of the selected polygons. This in no way nullifies the conclusions which can be drawn from the results, however, as the degree of error which was based on the data for 14 towns was greater than that found for the 18 towns.

to apply this result to the entire French territory; the sampling has been limited to the large cities and there is no reason to believe that the census of enterprises has been as incomplete in the rural areas as in the cities. Furthermore the high correlation between x and y might indicate that the converse is true; namely the denser the population of enterprises, the greater the number of omissions.

COMPLEMENTARY RESULTS

Although it may be interesting to know that in the large cities 20 to 27 per cent of the enterprises were overlooked in the general census, this information is nevertheless insufficient. One must ask oneself how and why the census has been incomplete. In particular, the first question that presents itself is the following: Have the enterprises been overlooked in the census because of their nature or their size? One might in fact think that principally the small enterprises have been omitted.

An answer to this question, at least with regard to the size of the enterprises, is furnished by Table 2, which compares the distribution according to the number of their salaried employees, of the 756,235

TABLE 2

DISTRIBUTION, ACCORDING TO THE NUMBER OF THEIR SALARIED EMPLOYEES, OF THE 756,235 ENTERPRISES INCLUDED IN THE CENSUS AND THE 1,623 ENTERPRISES NOT INCLUDED IN THE CENSUS BUT REVEALED BY THE SAMPLE

Number of salaried employees	Enterprises included in the census		Enterprises not included in the census but appearing in the sample	
	Number	Per cent of total	Number	Per cent of total
0	230,354	30	397	25
1	197,252	26	546	33
2 to 5	214,504	29	443	27
6 to 10	48,561	6	100	6
11 to 20	29,188	4	65	4
21 to 50	21,294	3	47	3
51 to 100	8,001	1	15	1
More than 100	7,081	1	10	1
Total	756,235	100	1,623	100

enterprises included in the census of March 1946 covering France's whole territory, with an analogous distribution of the 1,623 enterprises which the sample revealed were omitted from the census.

The two series of percentages show a very satisfactory agreement for all the enterprises having 2 or more salaried employees. For the two

other catgeories (0 and 1 salaried employee) the agreement is less satisfactory. It does not seem necessary, however, to attach much importance to this fact. First of all, the percentages resulting from the sample can be considered as only approximate, in view of the standard deviation of the sample. Moreover, the distribution of enterprises between these two categories is subject to some uncertainty. The enterprises comprising only two persons could be classified either as an enterprise consisting of two employers and consequently no salaried employee, or as one consisting of one employer and one salaried employee, according to the manner in which the informant interpreted the questionnaire. It is, in addition, quite possible that, particularly for tax purposes, an enterprise having 2 employers and no salaried employee had claimed having one employer and one employee, e.g. that the employer's wife had been erroneously recorded as a salaried assistant.

Under these circumstances, it seems wiser to put together the first two categories of Table 2. In this way, one obtains percentages of 56 per cent for the census and 58 per cent for the sample, which may be considered as sufficiently alike.

Thus it seems possible to conclude that, if the census of 1946 has been inadequate, it has been so for the enterprises of all sizes. According to the information gathered by the investigators of the sample, the fact that such a high percentage of enterprises was overlooked in the census was attributable to negligence on the part of the census enumerators.

For these reasons the Institut national de la Statistique has not published the results of the 1946 Census of industrial and commercial enterprises. Since then, others methods have been used to examine this subject. A permanent inventory of these enterprises has been established. It will give a more complete and exact picture of the industrial and commercial structure of France.

A PROCEDURE FOR OBJECTIVE RESPONDENT SELECTION WITHIN THE HOUSEHOLD*

LESLIE KISH
Survey Research Center, University of Michigan

In modern survey methods growing emphasis is placed on the objective selection of the sample. For surveys of the general population, increasing use is made of area sampling to obtain probability samples of households. Heretofore, scant attention has been given to the question of how to make an objective selection among the members of the household.

A procedure for selecting objectively one member of the household is given as used in four surveys of the adult population. Demographic data as found in the sample are compared with outside sources for available factors.

The Problem

TO OBTAIN random samples for surveys,[1] two basic conditions must be satisfied:

1. The sampling method must provide for a known probability of inclusion, other than zero, for every element of the population.
2. The method must be translated into a procedure that can and will be applied in practice.

Area sampling is gaining acceptance as a practical and reliable procedure for obtaining samples of households with known probabilities of selection.[2] In general practice these probabilities are equal either within specified strata or throughout the sample.

Given a sampling procedure—such as area sampling—which uses dwellings as units of sampling: when does the question of selection within households arise? There is no need of selection:

---

* Presented at the 107th Annual Meeting of the American Statistical Association, New York City, December 30, 1947.

[1] For discussion of sources of biases that may arise when personal judgment enters into the selection see: Hauser, Philip M. and Hansen, Morris H. "On Sampling in Market Surveys," *Journal of Marketing*, July 1944, pp. 26–31.

[2] King, A. J. and Jessen, R. J., "The Master Sample of Agriculture," *Journal of the American Statistical Association*, Vol. 40, No. 229, March, 1945, pp. 38–42.

Hansen, M. H. and Hauser, P. M., "Area Sampling—Some Principles of Sample Design," *Public Opinion Quarterly*, Summer, 1945, pp. 183–193.

Houseman, Earl E., "The Sample Design for a National Farm Survey by the Bureau of Agricultural Economics," *Journal of Farm Economics*, Vol. XXIV, No. 1, February, 1947, pp. 241–245.

"A Chapter in Population Sampling" by the Sampling Staff of the Census Bureau, 141 pages, Superintendent of Documents.

A. If the respondent is uniquely determined (as head of household or homemaker); or

B. If the household is the unit of analysis and any adult member can give equally valid information.[3]

However, if the household contains more than one member of the desired population, it may be regarded as a cluster of population elements.

One may decide to include in the sample every member of the population within the household.[4] However, this may be a statistically inefficient procedure in general (depending on the cost-variance relationship of the survey design), unless one of these three conditions holds:

A. Information about all members can be obtained from one of them.
B. There is seldom more than one member of the population in the household. For example: there is an average of 1.2 spending units per household.[5]
C. If the intra-class correlation within the household of the variables measured is of negligible size, or if it is negative.

These conditions are not met generally in surveys of the attitudes of the adult population. Furthermore, multiple interviews in one household may lead to undesirable interview situations. Hence, there is need for a procedure of selection that will translate a sample of the households into a sample of the adult population. There are no great theoretical difficulties, but a practical procedure must meet the demands of efficiency and of applicability. There are several alternatives, and the choice among them depends on a number of factors: the nature and distribution of the population, the objectives and design of the survey, and the available facilities. Under the latter we may distinguish the factors of cost, and of the nature and training of the field force.

The Conditions

It may be useful to describe very briefly the general sampling procedures used by the Survey Research Center.[6] A stratified random

[3] This assumption may be unjustified. See: Deming, W. Edwards, "On Errors in Surveys," *American Sociological Review*, August, 1944, p. 361.

[4] See, for example: Watson, Alfred N., "Respondent Pre-Selection within Sample Areas," No. 2 of Technical Series on Statistical Methods in Market Research, page 7, Research Department, Curtis Publishing Company, Philadelphia.

[5] Based on the Consumer Finances Surveys of the Federal Reserve Board, conducted by the Survey Research Center.

[6] For a fuller description see: Goodman, Roe, "Sampling for the 1947 Survey of Consumer Finances," *Journal of the American Statistical Association*, September, 1947, pp. 439–448.

selection of dwellings is obtained by means of area sampling. Small areas, segments, define the clusters of households within counties (or groups of adjacent counties) selected as primary sampling areas. In each of these areas there are trained interviewers who are employed on a part-time basis. Detailed sampling instructions are given to them in order to insure correct understanding and execution of the different procedures required by a variety of types of surveys.

The interviews are of the fixed question, free-answer kind, requiring generally from 30 minutes to one and one-half hours. Conducting the interviews in the home is believed to contribute toward a satisfactory interview situation. The interviewer calls at pre-designated dwellings; the respondent is then selected by a fixed procedure. Return calls are made to find the not-at-homes.

For each of a series of four surveys, conducted in June, August and December, 1946 and April, 1947 respectively, it was desired to obtain samples of about 600 interviews of the adult population of the Continental United States. The samples were distributed in 31 primary sampling areas: in four of the 12 largest metropolitan areas plus 27 scattered counties.

The population of the four surveys was limited to adults living in private households, excluding, because of interviewing difficulties and cost considerations, some segments of the population: armed forces; hospitals; religious, educational and penal institutions; trailer, logging and labor camps; and hotels and large rooming houses.

A procedure was devised for selecting one adult in each sample household. This procedure was favored over some alternatives (such as selecting every other adult found in sample households) for two reasons:

1. It was desired to take no more than one interview in any household, in order to obtain each interview before the respondent had a previous opportunity to discuss the questions. Furthermore, multiple interviews were thought to be statistically inefficient because of the expected correlation of attitudes within the household.
2. An interview in every sample household was desired in order to avoid making futile calls on dwellings without interviews.

With this procedure unbiased estimates may be obtained by giving each respondent as weight the number of adults in the household. Such differential weighting may in general increase the sampling error. However, in the present instance this increase is not great because of

the concentration in two-adult households. About 60 per cent of the households have two adults, about 10 per cent have more than three adults and about 1 per cent have more than five. Another result of this high concentration is that, unless the variable has a high correlation with the number of adults in the households, the difference between the weighted estimate and that in which each respondent has equal weight will be small; in case of small samples this difference may be negligible compared to the sampling error. For all attitudes thus tested in these studies the difference was inconsiderable compared with sampling error. However, in the comparison given in Table I there appear two differences of about two percentage points.

In order that the procedure may be applied and checked without great difficulty it is desirable to have a variable for ordering the members of the household; a variable than can be obtained by the interviewer objectively and easily. The age and sex of the members of the household were used for this purpose.

The Procedure

A "face sheet" is assigned by the sampling section to each sample dwelling unit; on each of these there are, in addition to the address of the dwelling, a form for listing the adult occupants, and a table of selection. At the time of the first contact with the household, the interviewer lists each adult separately on one of the six lines of the form; each is identified by entering in the first column his relationship to the head of the household (wife, son, brother, roomer, etc.). In the next two columns the interviewer records the sex and (if needed) the age of each adult. Following this the interviewer assigns a serial number to each adult: first the males are numbered in order of decreasing age, followed by the females in the same order. To assign these serial numbers it is necessary to obtain the ages of all adults only in that small portion of households in which there are two adults of the same sex and not connected by parent-child relationship.

Then the interviewer consults the table of selection; this table tells him the number of the adult to be interviewed. One of the six tables (A to F) is printed on each face sheet; each of the six tables is assigned to one-sixth of the sample addresses in a systematic manner.

Tables to select the person to interview:

Each of these tables was assigned to one-sixth of the sample addresses in the four surveys.

A	Number of adults in D.U.	1	2	3	4	5	6 or more
	Interview adult numbered:	1	1	3	2	5	1

B	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	2	1	3	4	2	
C	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	1	2	4	1	3	
D	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	2	3	1	2	4	
E	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	1	2	1	3	5	
F	Number of adults in D.U.	1	2	3	4	5	6	or more
	Interview adult numbered:	1	2	1	4	3	6	

The slightly changed order of selection in the tables given below has certain advantages over that of the above tables:

The low numbers of selection are concentrated in tables A, B and C therefore the procedure will yield a male respondent in a great majority of the addresses to which these tables have been assigned. Evening calls are necessary to find at home most of the male respondents; the interviewer may concentrate his evening calls at these addresses. Conversely the interviewer may use his time during the day by calling at addresses to which tables D, E and F are assigned, with an increased chance of success.

The proper fractional representation of each adult is approximated more closely without the necessity of printing many more forms; the chances of selection are exact for all adults in households with 1, 2, 3, 4, and 6 adults. Because numbers above six are disallowed, there are one or two adults in a thousand (generally young females) who are not represented; there is a "compensation" for them in the over-representation of number five in the households with five adults.

Relative frequency of Use	Table Number	If no. of adults in household is					
		1	2	3	4	5	6 or more
		Select adult numbered					
1/6	A	1	1	1	1	1	1
1/12	B1	1	1	1	1	2	2
1/12	B2	1	1	1	2	2	2
1/6	C	1	1	2	2	3	3
1/6	D	1	2	2	3	4	4
1/12	E1	1	2	3	3	3	5
1/12	E2	1	2	3	4	5	5
1/6	F	1	2	3	4	5	6

It may be noted that the procedure can be modified easily to select a constant proportion, say half or one-third, of the adults. In that case, of course, each adult would have the same chance of selection, regardless of the size of the household. In some of the households more than one interview would be taken, in some no interviews at all. The above tables may be modified readily to apply the changed procedure.

Results: Checks against outside sources

The distributions of the respondents were checked with outside sources whenever we could obtain or adapt reliable data for valid comparison. They are presented in Tables I and II.

TABLE I

COMPARISON OF RESPONDENTS IN SAMPLE WITH CHECK DATA

	Check Data	Total for 4 Surveys N =2372		Data from the Surveys of:			
		Weighted	Unweighted	June 1946 N =585	Aug. 1946 N =592	Dec 1946 N =570	April 1947 N =625
White	90.6	89.1	89.4	89	90	88	89
Native born	90.0	90.6	90.8	90	91	91	90
Age in years							
21–29	22.8	22.7	20.6	22	24	22	22
30–44	33.9	33.0	34.1	34	33	36	30
45–59	25.6	27.3	26.7	26	27	25	32
60 and over	17.7	16.1	17.8	17	15	17	15
NA		0.9	0.8	1	1	0	1
Education							
Not finished grammar	46.1	25.7	26.2	24	25	24	30
Finished grammar		19.1	19.3	17	17	22	20
Some high	17.4	17.1	17.4	18	19	18	14
Finished high	22.9	20.4	20.0	20	21	18	22
Some college	7.1	9.9	9.4	11	11	10	7
Finished college	5.0	7.0	6.9	9	6	7	6
NA	1.5	0.8	0.8	1	1	1	1

The check data are those deemed most nearly valid for comparison: the per cent white and the educational attainment are for April 1947 from Census Series P-20 No. 15, age is for July 1946 from Series P-S No. 19, and the per cent native born is the 1940 census figures for ages 16–64. In the data obtained from the survey results each respondent is weighted by the number of adults (from one to six) in his household. However, the column "unweighted", in which each respondent has unit weight, is included for purposes of comparison.

The appropriate sampling error (i.e. two standard errors) of the

estimates in Table I range from three to six percentage points for various items on each of the surveys, and from two to four percentage points for the four surveys combined. The estimates of color, nativity and age groups are in general agreement, with only one of the 30 estimates lying slightly beyond the given ranges.[7]

The data of the first three surveys showed an apparent upward bias in reports of college education. This was suspected to be a response bias to the single question on this topic. For the fourth survey three more questions were inserted to get more data on persons claiming 12 years of school or more; this involved additional work on only a third of the respondents. The fourth survey shows close agreement with the check data on college education, pointing to elimination of a response bias.[7]

Check of the sex ratio

TABLE II

PERCENTAGES OF MALES

Check Data: Census Estimate July, 1946	All Adults in Sample Households	Respondents in Sample Weighted by Number in Household				
		4 Surveys Combined	June 1946	August 1946	December 1946	April 1947
48.2	47.9	46.8	45	46	45	52

The Bureau of Census estimate is for the civilian non-institutional adult population for July, 1946 (series P-S No. 19). The second figure is an internal check, obtained by tabulating all adults listed as living in sample households.

Males appeared to be underrepresented among the respondents of the first three surveys. Although the difference was small, its presence in three surveys pointed to possible occasional deviation from rigorous procedure in the field. It may be noted that a 3 per cent bias in the sex ratio would require an unlikely 33 per cent sex difference in an attitude studied in order to produce a 1 per cent difference in survey results; a difference within the limits of accuracy of these surveys.

Tabulations of all adults in the sample households, as listed on the face sheet, gives close agreement with the check data. Continued investigation of the problem points to two sources of this small bias, both due to the fact that males are more difficult to find at home even with repeated call-backs: overrepresentation of males among the non-

[7] These statements are also borne out by results on three more national surveys in which the same procedure was used.

responses, and an occasional substitution on part of a few interviewers.

Two other small sources of error present in the first three surveys were eliminated by minor changes in the procedure.

Conclusion

The described procedure of selection within the household gave results that were satisfactory within the demands of the survey objectives. While there were occasional departures from correct procedure in the field, the procedure is such that extensive control over its field application, hence improvement, is feasible. It must be emphasized, however, that a practical sampling procedure is not an automatic device. For success it depends on a field force having both the training and the morale necessary for correct application.

BENEFICIARY STATISTICS UNDER THE OLD-AGE AND SURVIVORS INSURANCE PROGRAM AND SOME POSSIBLE DEMOGRAPHIC STUDIES BASED ON THESE DATA*

ROBERT J. MYERS
Chief Actuary, Social Security Administration, Washington, D. C.

The old-age and survivors insurance program, under the Social Security Act, is a nationwide social insurance system covering most employment in industry and commerce. It is possible that in the near future this coverage will be extended to virtually all employment in the country since almost all interested organizations are in favor of this. From the extensive scope of the present limited coverage and also of the anticipated, practically universal future one, there naturally will emerge over the years a very large number of beneficiaries, both retirement and survivor cases. This paper will first present a limited amount of background regarding the present program. There will then be set forth the sources and types of data now being presented or available; finally there will be described some demographic studies that have been made, as well as others which are possible from these data.

Coverage of the Program

THE EMPLOYMENT coverage of the present system includes substantially all wage and salary workers in industry and commerce except those engaged in railroad employment for whom a separate system has been set up. The major employment categories excluded from coverage are agricultural, domestic, governmental, non-profit, and railroad, as well as self-employment. Table 1 presents certain relevant data as to the extent of coverage under the system.

Benefit Provisions and Definitions

Monthly retirement benefits are available under the OASI program for "fully insured"[1] workers age 65 and over, and for their wives 65 and over and their dependent children under 18. These payments are not

* Presented at the 108th Annual Meeting of the American Statistical Association in Cleveland, Ohio, December 27–29, 1948.

[1] Roughly, defined as having had covered employment of a significant amount ($50 or more) in at least half of the calendar quarters since 1936, or attainment of age 21 if later, and before attainment of age 65 or death. Employment before age 21 and after 65 is credited toward meeting the requirement even though it does not serve to increase the requirement. In no case is there required more than 40 such quarters.

TABLE 1

COVERAGE OF THE OLD-AGE AND SURVIVORS INSURANCE PROGRAM

Description of Coverage	Number involved
1. Persons in covered employment during average week in 1947	34 million
2. Persons in covered employment some time or other during 1947	49 million
3. Persons, at end of 1947, who were at some time in covered employment during 11 years of the program	77 million
a. Number, in this category, who had earned sufficient wage credits to be insured	43 million
Average cumulative wage credits for this group	$11,000
b. Number with insufficient wage credits to be insured	34 million
Average cumulative wage credits for this group	$1,300
4. Equivalent amount of insurance in force under survivor provisions of the OASI system at end of 1947	$75 billion

made, however, if the individual earns $15 or more per month in covered employment. As a result, many individuals do not file claim upon attainment of age 65, since they continue at work. On the other hand, some persons file a retirement claim but return to work, either permanently or sporadically. Beneficiaries who actually receive monthly payments are said to be in "current payment status." When there are added those who have filed but are not currently receiving payment because of covered employment, the total is termed "benefits in force." At the end of 1948, there were 1.97 million persons age 65 and over who were fully insured, but only 53 per cent were in current payment status. Those who had filed a claim but were not in current payment status represented 9 per cent, while the remaining 38 per cent never had filed; thus 62 per cent of the benefits were in force.

Survivor benefits are available[2] to widows when they are age 65 or

[2] All survivor benefits are payable if the deceased worker was "fully insured"; in addition, even though "fully" insured status was not present, monthly benefits are payable to orphans and widowed mothers where "currently insured" status (roughly, covered employment of $50 or more in half of the quarters of the last 3 years) existed.

over, or when they have 1 or more dependent children under 18 in their care. Monthly benefits are available for children under 18 who are orphaned. In certain cases, there are monthly benefits for aged parents of deceased workers, if there was no surviving widow or child under 18. These survivor benefits are subject to the same concepts of current payment status and benefits in force.

Table 2 gives certain summary data on beneficiaries. It may be

TABLE 2

BENEFICIARIES OF THE OLD-AGE AND SURVIVORS INSURANCE PROGRAM IN CURRENT PAYMENT STATUS, DECEMBER 1948

Category	Number (in thousands)	Average Monthly Benefit*
A. Retirement		
Primary	1048	$25
Wife (supplementary)	321	13
Child (supplementary)	25	12
B. Survivor		
Widow, old-age	210	21
Widow, with children	142	21
Child	556	13
Parent	12	14
Grand Total	2314	†

* Rounded to the nearest dollar.

† Not computed because not significant.

noted that the 2.3 million beneficiaries are receiving payments at an annual rate of about $550 million. It has been recommended by the Social Security Administration and the Advisory Council of the Senate Finance Committee (Senate Document 208, 80th Congress, 2nd Session) that the benefit level be raised and additional beneficiary categories be added.

Another concept involved is the distinction between "initial" and "subsequent" entitlements as represented in the new claim awards of a particular year. An "initial" entitlement is the first claim on any particular case, such as a worker retiring, a wife's benefit claim filed when the husband retires, or a widow's benefit claim when she is over 65 or has dependent children when her husband dies. A "subsequent" entitlement is a claim which would not have resulted in monthly payments

at the time of the insured's death, as for instance a wife's benefit claim when she attains 65 after her husband retired or a widow's benefit claim when she attains 65 after her husband's death.

In contrast with temporary suspension of benefits, as for instance because of covered employment, there are a number of reasons for permanent termination of claims, such as death, remarriage of widows, and marriage or attainment of age 18 of child beneficiaries.

Sources of Data

Beneficiary data are published in a number of official publications of the Social Security Administration. The *Annual Report* carries only general summary data, such as number and amount of benefits in current payment status and total payments certified, allocated by state and also given for the entire country by category of beneficiary.

The *Social Security Yearbook* presents a large volume of statistics on a calendar-year basis. The most basic beneficiary data are the numbers and monthly amounts of benefits awarded, in current payment status, and in force by beneficiary type for various periods (also carried forward monthly in the *Social Security Bulletin*). There are given data on benefits suspended and on benefits terminated, classified according to reason. Data for the current year are shown for new awards of the year, benefits in force, and benefits in current payment status by age, sex, and race, as to both number and amount. Although the data are summarized into fairly broad age groups (generally quinquennial) because of desired economy in publication, the figures are actually available by single years of age. This, in many instances, is necessary for proper analysis, especially for retired persons where there may be a great difference between those who retire promptly at age 65 and those who defer retirement. Beneficiary data are also available to indicate in summary form the family composition of the beneficiary groups, both for new awards in the year and for benefits in current payment status at the end of the year.

The aggregate data on benefits are an actual count, whereas the other data giving details by age and family classification are based on a 20 per cent sample.

In addition to published periodical data, there is prepared each year a volume of substantive claims statistics which shows in considerable detail the claims awards (on a 20 per cent sample basis) by age of retired or deceased wage earner and type of benefits available.

Once each year there is obtained a count of the number of beneficiaries actually receiving payments in each county in the United

States. These data are subdivided only by type of benefit and not by age, sex, or other family characteristic and are used not only for publicity purposes, but also for contrast with the recipient-load under the various public assistance programs.

At irregular intervals, sample surveys on an interview basis are made to study various phases of beneficiary living conditions and resources. These surveys, although being based on the interview approach and subject to some limitations, are stratified samples of beneficiaries from the claims files. The results of these studies are published in the *Social Security Bulletin*, or in special releases. Also from time to time, there are issued various *Analytical Notes* which sometimes relate to special studies and tabulations dealing with different characteristics of beneficiaries. In addition, some of the *Actuarial Studies*, published intermittently, analyze beneficiary data and present future estimates thereof.

Adaptability of OASI Data for Demographic Studies

There are many demographic studies which may be made from the basic data collected under the OASI program. Some of these raw data now exist in tabulations and in analyses which are under way. In other cases, different types of coding and tabulating are necessary. For instance, in the past, deaths or other events such as remarriage have not been allocated by the date of occurrence, but rather classification was by date of administrative action, either original filing or final award. Complications, too, arise because of the several concepts described previously, namely, in force status versus current payment status and initial entitlement versus subsequent entitlement. These various complications can probably be resolved or taken into account by appropriate approximation or adjustment.

The various basic data which are tabulated and analyzed are necessary in the actual operation of the program. This is particularly the case in regard to actuarial cost estimates and other analyses for policy and planning, and also from the administrative standpoint in estimating work loads for the near future.

Mortality Studies

Certain crude mortality studies have been made. For primary beneficiaries (retired workers), it has been found that as contrasted with general population experience, the mortality has been somewhat higher for men (by roughly 5 to 15 per cent), but somewhat lower for women (by about 15 to 20 per cent).

It appears likely that mortality for male primary beneficiaries is

relatively unfavorable because with present high employment conditions, those who have actually retired, in most instances have done so only because of disability. This argument is further strengthened by the fact that those who retired promptly at age 65 and those who retired in 1940, the first year in which benefits were available, tended to have higher mortality than other retirants.

The mortality for female primary beneficiaries, despite the factors previously mentioned as relating to male mortality, may have been lower than normal because women who had been working near age 65 long enough to be insured might, on the whole, be superior lives. For both sexes, the death rates appear to have been independent of the size of benefits, which in turn reflects economic status.

Considering other types of beneficiaries, studies indicate that both wife and widow beneficiaries, all of whom are age 65 or over, had relatively low mortality as contrasted with general female mortality, namely, by an amount of 10 to 15 per cent relatively. The mortality among child beneficiaries, the vast majority of whom are paternal orphans, is about 10 per cent higher than would be expected from general population experience. At the same time, the mortality of the widowed mother beneficiaries is quite close to the general population experience.

A recent study made primarily to study remarriage experience has indicated that for white widowed mothers, the mortality was perhaps 5 per cent below that "expected," but for nonwhite widowed mothers there was lower mortality by about 15 to 20 per cent. The relatively favorable nonwhite mortality may occur because those nonwhite workers who are regularly in covered employment sufficiently to be insured probably represent those of above average economic status, and therefore the health conditions of their family would tend to be above average.

Unlike studies based on census data, the particular individual beneficiaries under OASI will be under observation over many years so that there is considerably more accuracy as to age reporting, both as to the living and as to the deaths. Eventually, there will be a large body of data which will yield information on that great "unknown" of mortality investigations, namely, the true experience at the older ages (say, 85 and over).

Also, from the OASI data it will be possible to examine mortality by marital status. Certain Census studies in the past have shown that married persons have somewhat lower than average mortality and that widowed persons have relatively high mortality. Preliminary studies

made from OASI data indicate that although married beneficiaries, both the retired workers and their spouses, have somewhat favorable mortality, there is no indication of very unfavorable mortality among widows, whether beyond age 65 or not.

By having the same group of lives continuously under observation, it will be possible to construct "generation mortality tables," which will give valuable indications as to the mortality of a specific group of individuals traced through from age 65, or slightly older, to eventual death. This contrasts with Census methods where it is only possible to study mortality by age for different groups of persons in a given year or short period of years; for instance, the experience at ages 80–84 will relate to different individuals than that at ages 65–69. The objection may here be raised as to amalgamating data which are not entirely homogeneous since they do not relate to a closed group of persons.

Remarriage Studies

The only available analyzed information on remarriage experience in the United States has been the American Remarriage Table which relates to a relatively small set of data arising from workmen's compensation claims under private insurance carriers.[3] Under the OASI system, there is a great body of experience since the number of new widows created each year is close to 100,000. For a study of this type it is necessary to consider both age and duration of widowhood.

One crude study that has been made in regard to the remarriage experience under the OASI system indicates that the actual remarriage rates are considerably higher than those of the American Remarriage Table, apparently by from 50 to 75 per cent. Several explanations seem likely. Since 1940 marriage rates have been quite high so that perhaps remarriage rates too have been high. Also the level of benefits under the OASI system is relatively low as compared to that under workmen's compensation claims so that there is less financial deterrent to remarriage.

At the lower end of the age scale, it is possible to obtain information as to mortality and marriage of children under 18 since these are reasons for termination of child's benefits.

Other Studies

If permanent and total disability benefits are added to the OASI program, as has been proposed, there will eventually be much data on

[3] Roeber, W. F. and Marshall, R. M., "An American Remarriage Table," *Proceedings of the Casualty Actuarial Society*, Vol. XIX.

disability incidence and termination rates. This has been a subject on which there is very little material in the United States other than the experience under life insurance policies and certain pension plans for governmental employees. Some students feel that neither of these sources of data nor the experience of foreign social insurance systems is indicative of the possible benefit experience which might arise for such a program in this country.

The beneficiary data under the OASI system can also yield information on family composition. Among the various data which could be obtained are the following: relative ages of husbands and wives, marital proportions, proportion of persons with children and numbers of such children, etc. In all instances, the analysis could be carried forward by age and race.

In the preceding discussion, the various demographic studies have been discussed on a nation-wide basis since this is the nature of the OASI program. However, the analyses could be subdivided by regions or even by states, but then there would be considerable administrative difficulty because the records are not set up to show existing residence combined with other demographic statistics. Likewise, any studies of migration of beneficiaries, although theoretically possible, would not be administratively feasible.

In addition, the analyses could in all instances be made by economic level since benefits payable are related to past covered wages; this would be especially valuable if coverage were extended to all employment, but under present limited coverage the results would not be too significant.

Although it would be highly desirable to make mortality and other investigations according to occupation, a subject on which there is available all too little information, it would appear that this would not be too feasible from the OASI data. For one thing, the records which are obtained through the working lifetime do not indicate occupation, although they do show the industries in which the worker was employed. However, because of the dynamic character of the labor market, many individuals shift frequently from one type of industry to another so that even if the lifetime data were available for each worker at time of retirement, there would be considerable difficulty not only theoretically, but even more important administratively, in classifying the "normal" industry.

As is apparent from the wide scope of the program, there is a vast amount of statistical data in the files which has not been tabulated or collected. The manpower limitations during the war years, of course,

had an appreciable effect on studies that might desirably have been made. Since the war, there have been both manpower and budgetary limitations.

As contrasted with the various mortality and other types of investigation that insurance companies do, the Social Security Administration has found it possible to do only a limited amount of actuarial and statistical investigations. In part, this might be explained by the fact that insurance companies must keep a close surveillance of their experience in order to be able to charge proper premium rates and to allocate dividends equitably.

On the other hand, under a governmental system, there are not such concepts of individual equity. Rather the benefit schedules are set by law and are payable accordingly, with only a general examination being required as to the immediate and long-range over-all financial balance. However, it is hoped that in the near future more extensive tabulation and collection of data will be possible so that some of the demographic analyses indicated previously may be made.

BY-PRODUCT DATA AND FORECASTING IN UNEMPLOYMENT INSURANCE*

NATHAN MORRISON
Division of Placement and Unemployment Insurance
State of New York

The unemployment insurance program has been operating in this country since the beginning of 1938. There have been many growing pains during these first eleven years. The operating problems involved in setting up this vast program have left little time thus far for careful consideration of the statistical data that are automatically produced in the course of its operations.

These data, now and in the future, can be a fruitful source of information on changes in economic activities. The comprehensive nature of these data, covering every city and county, giving information by detailed sub-industry, and yielding statistics for each week and month, makes it possible to test various theories of the business cycle, and may, as a result of careful analysis, reveal interrelationships among economic variables that are now unsuspected.

The data described in this paper are collected by all the state unemployment insurance agencies as part of their operations. However, some states publish only the statistics which are needed for administrative and operating purposes. Undoubtedly, more would be published if the existence of these valuable materials became generally known.

BY-PRODUCT DATA

WHAT DOES the unemployment insurance system offer in the way of by-product data? Essentially, it provides detailed geographical and industrial data on employment and payrolls in firms covered by the state unemployment insurance laws, and on unemployed persons claiming benefits.

The coverage of the unemployment insurance program is not yet complete. About 2 out of every 3 workers in the country are now included. In manufacturing, however, about 98 per cent are covered. The chief groups excluded are government employees, agricultural workers, self-employed persons, workers in some small firms (the state laws vary in this respect), employees of non-profit organizations, and domestic

* Presented at the 108th Annual Meeting of the American Statistical Association in Cleveland, Ohio, December 27, 1948.

workers. (Railroad workers have a separate unemployment insurance system.) Federal and state legislation to include practically all of these groups may be confidently expected within the next few years.

For many industries and geographic areas, the present laws represent substantially complete coverage. In terms of long-range studies, work can be planned on the safe assumption that full coverage will be reached in the near future.

A concrete example is perhaps the best way to indicate the data available and some of the ways these data can be used.

For every city and county in New York State, tabulations are available showing employment and payrolls by detailed industry group. The industry classification uses a 4-digit code that gives about 600 different sub-industry groups. The employment data show the number of workers during the mid-week of each month. The payrolls are on a quarterly basis. Since 1937, all employers with 4 or more employees (excluding the groups mentioned earlier) have been submitting such employment and payroll reports together with their tax contributions on a regular routine basis, as a normal activity of the unemployment insurance system. Actually, these reports contain more than the summary data on employment and payrolls. The name, social security number, and earnings of each individual worker are listed and submitted also each quarter. The reports are needed in order to collect the tax and to provide the individual earnings data used to determine benefit rights when an unemployed person files a claim for unemployment insurance.

These reports are mandatory under the law with severe penalties for failure to comply. The statistical data on employment and payrolls are collected as a by-product of necessary operations and are thus obtained at a minimum cost.

Another operating feature is the maintenance of lists, card files, and addressograph plate files, of all covered firms. Thus if a sample of firms is needed, the universe from which it is to be selected is readily available.

The usefulness of these data on establishments, employment and payrolls, by industry and geographic area, for market analysis, industry studies, labor market studies, and other related purposes is obvious. For study of changes in economic activity, Arthur F. Burns[1] has suggested the preparation of tables showing not only the total change in employment each month by industry, but also the number of firms increasing their work force and the number of firms dropping workers.

[1] Arthur F. Burns—"Twenty-sixth Annual Report of the National Bureau of Economic Research," June 1946, p 22.

Another type of analysis that may throw some light on economic changes is the study of the employment changes in each industry by size of firm. Some studies in certain New York industries have indicated that the largest and smallest firms have more stable employment than the medium-sized firms. New York State has also analyzed employment patterns in the millinery and dress industries in separate groups according to the price of the product, and in the canning industry by the crop being canned.

The collection of these employment and payroll data has involved problems of industrial classification and other troublesome details that are to be encountered in any large-scale undertaking. Such problems will always be with us. They need not concern the analysts who will use these data. One particular problem may be mentioned. At present, all employment and payroll reports submitted to the state agencies, and the monthly sample of establishment reports which are handled in cooperation by the Bureau of Labor Statistics and the state agencies, both cover the mid-week of the month (i.e., the week ending nearest the 15th of each month). The achievement of this uniformity is an obvious step, yet it required considerable effort and discussion.

The unemployment insurance agencies also provide data on unemployment as a by-product of their operations. The 48 states and Alaska, Hawaii, and the District of Columbia all have individual laws with varying eligibility requirements for benefits. These minor differences have been given exaggerated importance by some persons who are not familiar with the practical situation. Essentially, the unemployment insurance laws define an unemployed person as anyone who is not working, is seeking work, and is willing and able to take a job if offered one. This definition is simple and unambiguous, and is in agreement with the average person's idea of unemployment. An unemployed worker, in order to obtain insurance benefits, must report in person at the local unemployment insurance office nearest his residence, and must fill out forms and answer verbal questions which are designed to obtain information concerning his occupation, last employer, reason for loss of employment and his willingness and ability to take a job. This direct, personal contact with the unemployed worker is maintained, for he must report to the office once each week as long as he continues to be unemployed and seeking benefits.

Operating reports prepared weekly by each local office give the number of persons reporting each week by type of claim, i.e., first claim and continued claim. Central offices, which usually prepare and mail out the benefit checks, provide as part of their operations, additional data on

the previous industrial attachment and earnings of the claimants. The existence of these operating data offers the opportunity for many types of studies of the characteristics of the unemployed and of the incidence of unemployment by industry, area, and occupation at negligible expense. Two examples will illustrate some of these uses. Since 1940, New York has made a monthly sample study of benefit payments to determine the industrial distribution of the claimants on a current basis. More recently, there has been an increase of about 25 per cent in the number of claimants during the last two months of 1948. A study has been made of all first claims received during these two months to determine the industries which were laying off people.

The present incomplete coverage of the unemployment insurance laws is only of minor significance. In most of the cities in New York State, the persons now seeking benefits at the local insurance offices represent about 80 per cent of estimated total unemployment. Considering the detailed information available by area and industry, and the personal, repeated contacts with this large majority of the unemployed group, it is clear that the unemployment insurance program has provided as a by-product, a major advance in the field of unemployment statistics, far beyond anything available before 1935.

New York State has been working on a long-range study which other states have also been urged to undertake by the Social Security Administration.[2] Beginning with 1937, individual earnings records are available for each year to date for every person employed in covered establishments. These records show quarterly earnings with each employer. In addition, for those persons who were unemployed and filed claims for benefits, data are available for each year showing the number of weeks of compensable unemployment and the amount of benefits received. For many administrative and legislative purposes, and in order to supply data requested by the Social Security Administration, an annual study has been made in New York to determine the number of benefit claimants by industry, duration of unemployment, and benefit rate. By taking as a sample the approximately 10 per cent group of persons whose Social Security account numbers end in the block 2,000 to 2,999 (referred to as the "2,000 block"), and by entering the information on individual record cards which have space for 14 years of employment and unemployment data, it has been possible to obtain, as a

---

[2] A comparable long-range study of employment data available under the Old Age Insurance Program is described in "*The Continuous Work History Sample under Old Age and Survivors Insurance*" by Jacob Perlman and Benjamin Mandel, Social Security Bulletin, February 1944. New York's study has been developed in collaboration with the staff of the Social Security Administration.

	1937			1938			1939			1940			1941			1942			1943		
	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.	Q	Wages	Ind. and E. R. No.
Tot.																					
H.Q.																					

	1938-39 BEN. YR.	1939-40 BEN. YR.	1940-41 BEN. YR.	1941-42 BEN. YR.	1942-43 BEN. YR.	1943-44 BEN. YR.	1944-45 BEN. YR.
Did not apply							
Eligible claim							
Ineligible claim							
Local office							
Date of orig. claim							
Date of last check							
Time span of claim							
Spells of unempl.							
Benefit rate							
No. of checks							
Total benefits							

by-product of these required annual studies, a continuous record of the employment pattern of a large sample of New York workers over a period of many years. The detailed analysis of these materials has thus far been limited by recurring budget difficulties resulting from inadequate financing of unemployment insurance administration by Congress. However, the data are produced each year as by-products of the operations of the unemployment insurance system, and are being preserved for analysis whenever adequate staff becomes available. This long-range study of employment and unemployment by industry and area since 1937, which can be made in New York and the other states, should provide valuable data for both theoretical and practical work of students of the business cycle, business and market analysis, and economists and statisticians in general. (See figure 1.)

These materials have also been used to a considerable extent in forecasting work. This aspect can be touched on only briefly in this paper. The forecasting is done for two major purposes. Unemployment insurance benefits are paid out of a trust fund which in New York has totalled about $1 billion since 1945. The solvency of this trust fund is of course a major concern. Each year, all legislative proposals for changes in the benefit formula or in the taxing formula are analyzed and estimates of the effect on the solvency of the fund are prepared under various assumptions as to the level and movement of employment and payrolls during a 5 to 7 year period in the future. The effect of the end of the war was the subject of a comprehensive report which used all available materials on employment and unemployment by industry and area as the background. The second major purpose is to estimate claim loads for budgetary and planning use. Administrative funds are provided by grants from Congress. A year in advance of the beginning of each fiscal year, estimates of unemployment insurance claims loads expected during the fiscal year must be submitted by each state. In addition, forecasts are prepared for each semi-annual period about four months before the beginning of each period; and finally monthly forecasts are prepared for each local office about two weeks in advance in order that local office staffs may be shifted to meet the load most efficiently. New York's experience in eleven years of operation may be of some interest and perhaps can be presented in detail in another paper.

CONCEPTS OF EMPLOYMENT AND UNEMPLOYMENT

Early in 1948, the International Labour Office issued a report on "Employment, Unemployment, and Labour Force Statistics" which is

subtitled "A Study of Methods." This report was prepared for the Sixth International Conference of Labour Statisticians held in Montreal in August 1947.

This I.L.O. report comments in some detail on the methods being used in this country and in other nations to collect data on employment and unemployment and suggests a broad approach which is essential if the available statistical resources are to be used efficiently.

The collection of data on employment and unemployment in this country is being done for many purposes, and by many different agencies. However, only the unemployment insurance system provides detailed geographical and industrial information on both employment and unemployment entirely as a by-product of its routine operations. It is for this practical reason that it is likely that with universal coverage, the unemployment insurance program will be generally accepted as the primary source of this type of statistics. It should be noted here that as far back as 1937 the Committee on Government Statistics and Information Services which was jointly sponsored by the American Statistical Association and the Social Science Research Council, stated in its report on "Government Statistics,"[3] "the present voluntary system of payroll and employment reporting may be rendered obsolete with the advent of unemployment compensation laws, which promise to provide comprehensive returns as a by-product of administration."

There will of course continue to be need for some of the other methods now being used. Again the reason is a practical one. The problems that statisticians are concerned with in economic statistics in general and in studying changes in the level of employment in particular are complex. The problems will be solved, but as Professor Neyman remarked in 1937 in discussing the analysis of economic time series, "it will not be done today or tomorrow." It is important to recognize that all available tools and resources are needed if significant progress is to be made in this field of research. In a paper published 21 years ago, Professor Harold Hotelling[4] pointed out that the Law of Gravitation could not have been established as it was by Newton if the sun were smaller and closer in size to some of the planets. If the sun were smaller, the family of planets would exhibit some of the complex movements of democratic societies and statistical methods not known in the 17th century would have been necessary to determine the Law of Gravitation.

The problems of employment and unemployment are important to

---

[3] "Government Statistics"—Bulletin 26, Social Science Research Council, April 1937, p. 92.

[4] Harold Hotelling—"Differential Equations Subject to Error and Population Estimates," *Journal of the American Statistical Association*. Sept. 1927, p. 287.

many groups and agencies in this country. The Joint Congressional Committee on the Economic Report in its recent report on "Current Gaps in Our Statistical Knowledge" refers to the need for extension of employment and unemployment statistics. The Committee states:

> "Such statistics are among the most important indicators of the economic situation. In this area, the Census Bureau's Current Population Survey is particularly valuable. The chief weakness of this monthly enumerative survey is that it fails to show employment and unemployment on a geographical basis, and that the sample is too small to provide reliable data on occupational and other characteristics of the unemployed. . . . The need for geographic detail can be met in part by annual surveys of individual metropolitan areas . . . Annual surveys of employment and unemployment for the major urban areas do not provide sufficiently current data in a period of rapid change. For this purpose primary emphasis must be placed on employment data reported by employers. State data of this sort are compiled by state employment security agencies and state labor departments in cooperation with the Bureau of Labor Statistics which also compiles the corresponding national series, but area series are maintained by only a few state agencies."

There is no reference at all to the unemployment insurance data available on claims filed by unemployed persons which every state agency collects as a by-product of its operations and which are actually reported to the Social Security Administration in Washington each month for each local office in every state.

The reference to the employment data reported for each month by all covered employers fails to emphasize the important fact that the data are available as a by-product of operations, and that the additional money needed for analysis and publication is but a tiny fraction of the cost of obtaining such detailed data on a current basis by any agency not conducting a vast program like the unemployment insurance system.

It is not possible in this paper to discuss the Census Bureau Household Sample in any detail. Obtaining information by visiting a sample of households has some advantages and some important defects. This tool of research can be a useful supplement, at present, to the unemployment insurance by-product statistics. When universal coverage is attained, there will be little need for a household sample, except perhaps for some special purposes.

There is an important point concerning the definitions of employment and unemployment. The Census Bureau uses a definition of unemployment which is not in accord with the concept used in the state laws described above. The definition suggested by the International Labour Office report is similar to the one used by the unemployment

insurance laws, namely that "the unemployed should include all persons seeking work on a given day who are not employed but are able to take a job if offered one."

The I.L.O. report also comments on the Census Bureau insistence on including as unemployed only those who have been without work for at least the Monday through Saturday period in the enumeration week. The I.L.O. report estimates that unemployment was thus understated about 7 per cent during the period June 1941–May 1942.

It would be a major step forward if the Census Bureau would bring its concepts and definitions more in line with the concept of unemployment which underlies the laws of the 51 unemployment insurance agencies. Whatever technical difficulties stand in the way, they must be overcome if the Census Bureau work in this field is to play its proper role as a supplement to the by-product data.

The I.L.O. report has some pertinent comments on the use of unemployment insurance statistics as a measure of unemployment.The limitations it mentions are of some significance at present. However, with universal coverage and a probable extension of duration of benefits to 26 weeks in every state, these limitations will no longer be significant.

One quotation from the I.L.O. report will serve to summarize some of the views expressed in this paper:

> "Social insurance statistics of unemployment are in a very real sense cost free, being by-products of the operation of a system installed for other than statistical purposes. It therefore becomes possible to expand statistical coverage and derive additional estimates (as well as to secure the basic estimates) at a remarkably low cost assignable to statistical purposes. Furthermore, the continuing contact of a social insurance system with individual workers makes it possible to conduct a variety of special studies on unemployment problems at low cost and with little inconvenience to the employee. The possibilities of using social insurance data for the study of unemployment problems have only begun to be explored."

The rich mine of statistical data available as a by-product of our unemployment insurance system if carefully analyzed may very well lead to significant progress in economic theory and to a better understanding of the movements of the business cycle. Analysts and research workers in this field should examine the data already available, and should encourage the analysis and publication of data which are being collected in all the states, but which, in some states, may merely be stored away or filed for lack of funds and outside interest.

STATISTICAL REQUIREMENTS FOR ECONOMIC MOBILIZATION

RALPH J. WATKINS
Office of Plans and Programs
National Security Resources Board

IT IS IMPORTANT at the outset to make clear the distinction between mobilization and mobilization planning. In dictionary terms, mobilization is defined as the "act of assembling, equipping, and preparing military and naval forces for active hostilities." It is scarcely necessary to say that the American nation is not engaged in mobilization; nor are we preparing for mobilization. Mobilization planning is quite another matter, as I shall try to show presently.

The National Security Act of 1947 established the National Security Resources Board and assigned to that Board the statutory function of advising the president concerning the coordination of military, industrial, and civilian mobilization, including advice to the President on certain specific matters having to do with effective mobilization of resources in the event of war and with certain economic readiness measures against the contingency of war. Thus these functions make of the National Security Resources Board an economic mobilization planning agency set up to advise the President. It is important to note that the Board has no operating functions in the governmental sense of the term; its duty is to advise the President.

Economic mobilization planning may be defined as the process of estimating the requirements or needs of war; of appraising the resources or means which would be available for meeting those needs; of measuring deficiencies revealed by the comparison of needs with means; and of determining the steps necessary to balance needs with means—all to the end that there may be available well-articulated plans for mobilizing the resources of the nation in the event of war. It is important to stress the words "in the event of war." We are not engaged in planning "for war"; rather, we are planning against the contingency of war.

The philosophy of mobilization planning, both military and economic, rests on the premise that a state of preparedness is one of the means of lessening the likelihood of an aggressive attack against the nation and at the same time one of the means of increasing the likelihood of winning a war, if the nation is forced into war. In the un-

certain world in which we live, we can with prudence do no less than to take appropriate steps to improve our economic readiness position against the contingency of war and to lay plans for the rapid and effective mobilization of our resources in the event of war.

It will be recognized at once that the definition of economic mobilization planning which I have outlined is one which in many fields—not all fields by any means—lends itself to translation in statistical terms.

For example, one of the specific statutory functions assigned to the National Security Resources Board is to advise the President on "the relationship between potential supplies of, and potential requirements for, manpower, resources, and productive facilities in time of war." We interpret that provision to mean that this advice must be given in time of peace against the contingency of war, as well as in time of war. The basic and fundamental tool in a program of economic mobilization planning consists of detailed balance sheets of resources and requirements—above all for key raw materials but also for major end products, critical components, fuels, electric power, transportation, and manpower.

In time of peace, such balance sheets provide the factual basis for advice to the President from time to time as to the steps that should be taken to strengthen our economic readiness position or to lessen our economic vulnerability against the contingency of war. In time of war, this balance sheet process would provide the factual basis for operations by the war agencies.

What are we doing to meet this fundamental need for resources—requirements balance sheets? We have under way three basic programs:

1st. Quick and approximate estimates of requirements to test the economic feasibility of strategic plans.

2nd. Balance sheets for Fiscal 1950 of key resources and security program requirements.

3rd. Detailed estimates of resources and mobilization requirements on a systematic and continuing basis.

These three programs will be outlined briefly.

First—*Quick and approximate estimates of mobilization requirements to test the economic feasibility of strategic plans.* The Munitions Board in the National Military Establishment is now developing quick estimates—or flash estimates as they call them—of military requirements under the assumptions outlined in the strategic plans. These will embrace the needs of the Military Establishment for the key materials of steel, copper, aluminum, and petroleum; for construction; for many key

items of military equipment; and for manpower. When completed, these estimates will be submitted to the National Security Resources Board. Concurrently with this work, the National Security Resources Board is securing comparable quick estimates of mobilization requirements from the other security agencies—the Atomic Energy Commission and the United States Maritime Commission. In cooperation with the Departments of Commerce, Interior, State, and Labor, we are also developing quick statistical estimates of the mobilization requirements of the civilian economy. These last named civilian mobilization needs are being estimated initially on the basis of the civilian economy as it existed at the peak of our war effort in World War II, with, of course, appropriate adjustments for changes in population, changes in inventories, and changes in patterns of consumption. We will then add these estimates to those supplied by the Munitions Board to secure a measure of estimated *total* mobilization requirements.

We will next compare these mobilization needs with the estimates of available resources, which are being prepared also in cooperation with the Departments of Commerce, Interior, State, and Labor. On the basis of that balancing of mobilization needs against means, we will be able to say to the National Military Establishment whether the strategic plans can be encompassed within the limits of our resources, and if not, why, and in what particulars.

Second—*Balance sheets for Fiscal 1950 of key resources and security program requirements.* At the request of the President, the Board has undertaken, with the cooperation of 21 federal departments and agencies, to appraise the resources available to the nation in relation to the requirements in Fiscal 1950 of the several national security programs—current and anticipated. That survey will include a wide range of balance sheets covering the major strategic and critical materials, fuels, power, transportation, manpower, and the key manufacturing industries of the nation. These balance sheets will serve as the basis for advice to the President on questions of national security policy. For example, they will indicate: (1) to what extent our physical resources may set limits to the national security programs; (2) to what extent these programs can be accomplished without controls; (3) to what extent controls would be required to assure completion of these programs; or (4) to what extent controls would be required to protect the economy from price spirals in the raw material markets. Likewise, these balance sheets will afford a factual basis for economic readiness measures aimed at increasing our resources which the Board may see fit to recommend to the President.

Third—*Detailed estimates of resources and mobilization requirements*

on a systematic and continuing basis. Coming back to the assumption that the strategic plans meet the tests of economic feasibility which I have described: On the basis of the assumptions outlined in those plans, the National Security Resources Board must assemble from the several federal departments and agencies on a continuing and systematic basis estimates of key and major resources and key and major mobilization requirements, both military and civilian. The extent to which this statistical planning process can be projected currently remains to be determined. In consequence, I can merely outline in general terms my own views. The Munitions Board, with the cooperation of the Service, will estimate detailed military mobilization requirements for submission to the National Security Resources Board. The Board must look to the appropriate civilian deparments and agencies for the estimation of civilian mobilization requirements and likewise for estimates of available resources, as their continuing areas of responsibility in the task of economic mobilization planning.

The staff of the National Security Resources Board is now engaged in the preparation of requirements manuals. These manuals will outline the necessary assumptions, methods, procedures, and forms. Out of this process could be developed balance sheets covering major and critical needs. For their review, an Interdepartmental Committee on Program Balance, with appropriate subcommittees, may be necessary for critical screening and review, in the manner followed by the Requirements Committee in the War Production Board during World War II. The objective would be to plan for a balanced program for military and civilian requirements, against which resource allocations could be planned.

We realize, of course, that the problem is so great and so complex that we cannot hope, in time of peace, to do more than to design the machinery; to train a nucleus personnel within the National Security Resources Board and other agencies; and through peacetime operations produce, as the factual basis for mobilization planning decisions, balance sheets covering only key and major end-products and components, and their translation into requirements for critical raw materials, fuel, power, transportation, and manpower. Through this continuing balance sheet operation we would serve four basic mobilization planning purposes:

1st. We would provide systematic measurement of the relationship between resources and mobilization requirements and thus provide a factual foundation for advice to the President on economic readiness measures.

2nd. We would develop procedures and techniques which in the

event of war would be available for determining in detail the hundreds of resources-requirements balance sheets on which the war agencies could plan their operations.

3rd. We would develop procedures and interdepartmental administrative machinery for analyzing and integrating these balance sheets to achieve program balance.

4th. We would train a skeleton staff within NSRB and in other agencies of government which could quickly be expanded to conduct this vital operation.

It may properly be asked where the statistical records can be found for the economic mobilization planning which I have described. The general answer must be that in time of peace these must come from existing records and established channels for the statistical reporting of economic intelligence. Special mention, however, should be made of the valuable statistical materials to be found in the reports and records of the War Production Board and other war agencies. Of particular value are the records on bills of materials for translating end-product and component requirements into their raw material content. It may be added also that the Armed Services are adding to our bills-of-materials records through special provisions in procurement contracts.

Let us now consider briefly the nature of statistical requirements in time of war. Our experience in two world wars has shown clearly that effective prosecution of a major war is not possible without resort to numerous economic controls. National policy then demands that the sum total of business operations accomplish specific and well-defined objectives. To meet these objectives, government is forced into day-to-day relations with business units. Government administrators, as a result, must collect data showing a balanced factual picture of the economy. Business management requires a continuing flow of accurate and detailed information to make the kinds of decisions needed to conform to government regulations and to effect the maximum contribution of the concern's resources to the national objectives.

Statistics are needed to devise these war-time controls and to serve in their enforcement. To determine the nature and the extent of controls of manpower, for example, we need detailed statistical information on geographic distribution of employment, on the availability or non-availability of specific skills for specific industries in specific areas. Likewise, industrial controls cannot be established and effectively enforced without detailed information on needs, on usage, and on inventories. This information must be far more detailed and specific than that ordinarily available to the federal government. This information

must give us facts not only on the quantities of general types of commodities and products but on their varieties and shapes. It must give accurate indication of the physical location of inventories, their ownership, and their intended use. Similar detailed information is needed on capacities, production plans, and schedules, on storage and transportation, and on a multiplicity of other activities and needs. The collection of such information must be centrally planned. Its analysis and interpretation must be centrally coordinated.

Once initial controls have been established, detailed statistics are required as tools for the daily tasks of administering those controls. It is then that government finds itself in business on a large scale, buying and selling enormous quantities of materials and equipment. Conducting business on such a scale calls for managerial statistics which are intimately related to the accounting records from which such statistical information is derived. This type of statistics is in daily use by the management of private enterprises, but in time of peace the collection of such statistics is generally beyond the province of the federal government. In time of war, however, the federal government directs a large share of the economic activities of the nation, closely controls many of them, and must inject itself in many day-to-day transactions between sellers and buyers. Statistics required by the government to perform these functions naturally are not peace-time voluntary reports but the mandatory collections of a mass of detailed facts the reporting of which is justified by the emergency of war.

Perhaps the simplest way of indicating the magnitude of the problem of war-time control statistics is to note that the *Catalog of War Production Board Reporting and Application Forms* required nine large-sized volumes aggregating more than 3,000 pages merely to reproduce the most useful 1,200 of the 4,400 questionnaires developed and used by that agency; and an additional volume to list alphabetically the 5,000-odd commodities reported on in these 1,200 forms.

As stated in the General Introduction to that *Catalog:* "To mobilize our production resources for war, to insure that the products of mines, forests, factories, and chemical laboratories were utilized efficiently, to integrate materials and components into the greatest possible volume of finished products and to channel the distribution of finished products to the military and export agencies and to the domestic civilian population in the manner most directly related to winning the war, the WPB assembled a quantity of information of unprecedented magnitude and detail. The statistics collected included data on basic materials, semifabricated materials, components, subassemblies, and finished end

items. They also provided related information on the purpose for which the products were used and the kinds of ultimate consumers who used them."

It would, of course, be dangerous to assume that the statistical activities of the emergency agencies in World War II represented models of efficiency. Quite to the contrary, students of the problem can cite numerous examples of bad planning of statistical reporting and the collection of voluminous facts which sometimes could not even be tabulated fast enough to serve as guides to administrative action in a rapidly changing situation. It is, therefore, incumbent on statisticians engaged in mobilization planning to look with a critical eye on the statistical reporting of World War II; to seek earnestly to learn from that experience through selection of the good patterns and discarding of the wasteful and unnecessarily complicated patterns; and to strive for simplicity in designing the statistical controls for a possible future emergency.

THE WAR PRODUCTION BOARD'S STATISTICAL REPORTING EXPERIENCE, V AND VI

David Novick
University of Puerto Rico
AND
George A. Steiner
University of Illinois

This is the last of a series of four articles, in six parts, concerned with the statistical reporting experience of the WPB and its predecessor agencies. In this article there is presented an analysis of a few of the outstanding tabulation problems of the WPB and its principal tabulating agent, the Bureau of the Census. In addition, the series is concluded with a statement of a few outstanding lessons which the WPB's statistical experience has revealed to be strategic in the establishment of a suitable emergency industrial statistical reporting program.

PART V

TABULATING DATA

INTRODUCTION

The success with which the WPB was able to acquire accurate data and to compile them quickly into meaningful summaries determined in a very real way the effectiveness of basic policy decisions and administrative control actions. The tabulation of statistics should be an elementary process. Under wartime pressures and data demands, however, there were hidden in this apparently simple process many dangers. The greater the volume of data to be tabulated and the closer it related to operations the greater these dangers became. Improper arrangement of even the most minor details in final summaries many times invalidated an entire tabulation and the details from which it was compiled. Such failures produced a lack of confidence in a series, making the figures ineffectual for the purposes designed.

Principles and procedures used to aggregate statistical series into summaries are basically the same for both great masses of data and figures in small volume. But the WPB came to learn that problems incident to massive statistical compilations, particularly when they were to be used for administrative rather than purely statistical purposes, were of real significance to its effective operation. The main thesis of this article is to study a few outstanding tabulation problems facing

the WPB in its attempt to aggregate masses of data for managerial functions, to show how they were solved, and to note the impact on WPB management.

Wartime Complication of the Tabulating Process

The wartime tabulation process was complicated by three elements: (1) pressures of time, (2) the volume and complexity of numbers demanding addition, and (3) special wartime tabulation needs which required adaptation and often important modification of peacetime mass tabulation techniques. These forces became the more difficult to control because of severe limitations of time and of large-scale tabulation facilities and experience.

Wartime operations pressed on the time cycle of statistical processing from every direction. The result was an acceleration of all steps in tabulation procedure. The greedy demands of wartime operations were never satisfied and all energies were constantly directed towards effecting a further tightening of the tabulation time schedule.

The avalanche of statistics collected by the WPB severely strained available resources for aggregating them and gave rise to a series of problems. Mechanical and personnel resources were at times strained to the breaking point.

Wartime operational needs injected new problems for both industry and government into peacetime tabulation techniques. For industry they necessitated the establishment of new records, finer detail in records, an unremitting flow of information, and quick filing of reports. The result naturally was reporting inconsistency, incompletion of forms, failure to use the specified measures and outright clerical errors of serious proportions. Large-scale tabulation methods faced the problems of quickly unifying this extraordinary mass of diverse information and aggregating it in a way often different from that specified on the basis of past experience. The penetrating eye of war sought much more detail in summaries and greater accuracy than was required under ordinary circumstances. These were not the result of unreasonable expectations of the novice; they sprang from the urgencies of immediate administrative action which probed deeply into the heart of industrial production and distribution.

Two fundamental results grew out of these considerations. First, the care born of leisure could not be devoted to the methods and devices by which basic data requests were formulated and answered. Second, time was not available to permit the careful planning of all aspects of the mass tabulation procedure. Even so, many of the new

problems were matters which could not be resolved by analysis of peacetime experience but had to be answered by trial-and-error methods. The net result was the development of one more strain on the entire administrative process which in turn multiplied and hardened still further the compulsions of time.

The Production Requirements Plan Tabulations[1]

In a very real sense, form PD-25A, the operating instrument for the PRP, embodied most of the more serious tabulation problems encountered by the WPB and its most important tabulating agent, the Bureau of the Census. For this reason it is informative to study in detail the process by which form PD-25A was tabulated, especially the nature and technique of its product coding. For this purpose we shall choose the tabulation for the fourth quarter 1942 PD-25A upon which were reported second quarter 1942 actual metal use and fourth quarter 1942 critical material requirements. Although this may be considered typical it should be mentioned that various changes in the form itself; manufacturer and WPB experience in processing; and changes in the variety of final tabulation requirements—all resulted in slightly different tabulation techniques and problems for each PD-25A quarterly tabulation.

The first step in the tabulation of PD-25A was sorting schedules to facilitate coding, review, and tabulation. The most important sorting was for size. In view of the fact that it was impossible to tabulate within the time available all PD-25A schedules received (about 35,000), smaller plant schedules were sorted out and marked for exclusion from tabulation. If a company reported less than 30 tons of total steel or 10,000 pounds of all other items together its schedule was rejected for purpose of tabulation. The reason for this treatment grew out of studies made regarding the extent of metal concentration among manufacturers and the fact that it would have been impossible to meet a tight time schedule if the tabulation included the thousands of such small cases. On the basis of PD-275 returns it was calculated that the small cases rejected and the thousands of small plants not required to file PD-25A, consumed no more than five per cent of the metal used by all plants meeting PD-25A reporting standards, as defined in priorities Regulation No. 11. The small cases which were rejected consumed less than one-half of one per cent of the total material used by all industry.

[1] For a summary of the operation of the Production Requirements Plan, see the first article in this series, Volume 43, Number 242, June 1948, p. 220.

The next step in tabulation related to product coding of the schedules to be included in the tabulation. It was in this area that the major problems arose in PD-25A tabulations and their use.

The purpose of a product code is to assemble products which are the same, similar, or closely related, into fairly homogeneous groups by assigning the same number to all items included within the same group. These numerical symbols identify the product groups throughout the process of tabulation thereby insuring final summarization for comparable products. In the process of summarization, figures representing material consumption by producers of edge tools, for example, are meaningful only if it is reasonably certain that the figures include all reports for producers of certain specified edge tools. Uniformity of concept of product groups becomes indispensable when condensed reports by commodity groups are required without listing individual items. Under these conditions the choice is a simple one: uniform product classification or completely useless tabulation.

Various government agencies, particularly the Bureau of the Census, for many years had devoted considerable attention to the proper classification of industrial data. But this work, as discussed in Article III, Part IV, was not suitable for wartime use and as a result new classifications had to be developed. The fruition of this work, it was previously pointed out, resulted in the product codes used to tabulate PD-275 and PD-25A. Under the code worked out for tabulating these forms, it was possible to classify production data with the specific product to which they related and to provide summary information which the WPB could use for product administration.

The heart of the difficult coding problem of the PD-25A tabulation lay in the desire of WPB Divisions to obtain as fine product details as possible. The basic limitation on product group expansion, however, lay in the extent to which manufacturers themselves reported detailed data by individual products. PD-25A applications were developed by a manufacturer upon the smallest operational unit which his experience led him to establish. For the most part these units reflected common experience in terms of fabricating facilities, labor problems, processes, materials used, or products made. As a result, there existed in each industry a substantial group of closely related inventory control units, even though the ultimate use of the output may have been as dissimilar as vitreous enamel kitchen utensils and vitreous enamel hospital utensils or Army and Navy mess equipment. Summarization of PD-25A data for policy and administration was limited by the extent to which current industry records could relate material use, inventory,

consumption, and requirements to detailed products. Efforts to depart very far from the reports themselves could easily invalidate the final tabulations.

Hindsight leads to the conclusion, however, that a substantial educational program explaining the needs of the PRP and PD-25A might have provided the needed inspiration for manufacturers to report finer product details. The PD-25A, covering requirements for the first quarter 1943, contained such a statement of the product detail desired. The results permit the conclusion that had such information been given industry from the time of the first PD-275 the end result might have been far finer product classes than were finally developed.

Each PRP schedule contained three product codes: the minor, the secondary, and the major product code. Each code was based on 230 product groups (later expanded to 440) and was determined by the description of the products reported in Section B of the PD-25A application, shown in the accompanying illustration. The minor code represented the finest product detail reported and made possible a tabulation of dollar shipments for each product, the summarization of which was completely related to a homogeneous class of products. The secondary product code, the most significant of all, represented the most detailed breakdown of material data in relation to product groups. It was this code, into which the vital data for WPB material distribution operations were tabulated, that set the patterns of over-all metal authorization for groups of products, and it was this code which created most of the controversy surrounding PD-25A tabulation. Its exact composition, therefore, deserves further comment.

Product shipment information was reported in Section A of PD-25A. Total detailed metal consumption, inventory, and requirements for all products produced from materials included in the smallest inventory unit of the plant were reported in Section E, Part 1, and material usage and requirements for specific products were presented in Section E, Part 2.

If a manufacturer reported in Section E, Part 2, information for each product specified in Section A, the minor and the secondary codes would have been the same. But if, as was more common, a manufacturer reported fewer products in Section E, Part 2, than in Section A, the two would be different. This happened because manufacturers reported product details in Section B which could not be related to materials in Section E, Part 2. A producer of circuit breakers, for example, might report shipments for a variety of types in Section B but lump main types in Section E, part 2, simply because materials used

Attach a catalogue or other descriptive literature, if not now on file

DO NOT USE

Form PD-25A
(7-8-42)

UNITED STATES OF AMERICA
WAR PRODUCTION BOARD
WASHINGTON, D. C.

APPLICATION FOR PRIORITY ASSISTANCE UNDER PRODUCTION REQUIREMENTS PLAN

For Calendar Quarter October–December 1942

Submit six copies

Read instructions in Section A carefully before filling out Form

Serial No.	Name and Address of Company	Prior.
		No.
		Coded
		Entered
		Tab

Plant to Which This Report Applies	Location of Plant (Street and number city State)

If Entire Plant is Not Covered by This Report State Department or Operating Unit Covered

SECTION B *Part 1*—Actual Shipments Show below the number of units and dollar volume of shipments (including transfers to other divisions departments or plants) of each of the classes of products shown in column (1) Please read the instructions in Section A for guidance in grouping the products you manufacture Entries must appear in columns (3)–(7) and (9)–(10) Where the answer is None enter the word None.

DO NOT USE	Description of Classes of Products (1)	DO NOT USE (2)	Actual Shipments During Calendar Quarter April–June 1942: Total Dollar Value (Omit Cents) Analyzed by Preference Ratings: Total (3)	Allocated or specifically directed (4)	AA (5)	A 1-a (6)	A 1 b through A 1 k (7)	DO NOT USE (8)	A 2 through A 8 (9)	Other and nonrated (10)	DO NOT USE (11)
	Total										

SECTION B *Part 2*—Estimated Shipments. Show below in column (3) total estimated shipments for the calendar quarter October–December 1942 Enter in columns (4) (10) only *rated* orders now on the books scheduled for shipment during the quarter Enter in column (11) the balance of estimated shipments for which no rated orders are on hand. See instructions Section A. Entries must appear in columns (3)–(11) Where the answer is None enter the word None

DO NOT USE	Description of Classes of Products (1)	DO NOT USE (2)	Estimated Shipments During Calendar Quarter October December 1942: Total Dollar Value (Omit Cents) Analyzed by Preference Ratings: Total (3)	Allocated and AAA (4)	AA-1 (5)	AA-2 (6)	AA-3 (7)	AA-4 (8)	A 1-a through A 1 k (9)	A 2 through A 10 (10)	Other and nonrated (11)
	Total										

SECTION B *Part 3*—Total Unfilled Rated Orders: Show here the total dollar value of unfilled *rated* orders as of the most recent date available.

Total, $ .... as of .... (Give date)

SECTION C. } Information called for by these sections in previous PD-25A
SECTION D. } forms is not required.

Page 1 of 10—Applicant's Work Sheet

(Name and location of plant) (Department or operating unit)

SECTION E, *Part 1.*—Report all materials shown on Materials List No. 1, Revised, used in the manufacture of products listed in Section B, or for maintenance and repair. When materials are used in the forms shown on the portion of the Materials List printed below on pages 2, 3, and 4, enter quantities opposite the appropriate items. Quantities must be reported in the units indicated in column (3). Enter other items on the Materials List on blank page following the printed portion of the list, using the units specified in the Materials List. Even though you do not keep your records in these units, please convert into the units specified. Report quantities in columns (4) through (8), except for chemicals (see Section A, par. 32).

Item No.	Description of Materials (1)	Repeat Item No.* (2)	Unit of Measure (3)	Statement for Quarter April-June 1942: Received during quarter (4)	Put into production (used during quarter) (5)	Inventory on June 30, 1942 (6)	Estimated Inventory on September 30, 1942 (Optional) (7)	Statement for Quarter October-December 1942: To be put into production (anticipated requirements) (8)	DO NOT USE (9)	DO NOT USE (10)
	IRON AND STEEL			Quantity	Quantity	Quantity	Quantity	Quantity		
	Iron:									
100	Pig iron		Short tons.							
	Castings (before machining):									
101	Gray iron		Short tons.							
102	Malleable		Short tons.							
106	Pipe		Short tons.							
104	Bars and rods, wrought		Short tons.							
105	Pipe and tubes, wrought		Short tons.							
107	Plates and all other, wrought		Short tons.							
108	Cast iron scrap—inventory figures only		Short tons.							
109	Steel scrap—inventory figures only		Short tons.							
	Carbon Steel:									
110	Ingots, blooms, billets (except shell steel), slabs, sheet bars, and tin bars		Short tons							
111	Billets, shell steel		Short tons.							
112	Tube rounds		Short tons.							
113	Structural shapes and piling		Short tons.							
114	Plates, including skelp in plate sizes		Short tons.							
117	Skelp (except in plate sizes)		Short tons.							
118	Rails, over 60 pounds		Short tons.							
119	Rails, all other		Short tons.							
120	Tie plates and track accessories		Short tons.							
[illegible]	[illegible], except shell steel and con- [illegible]		Short tons.							
			Short tons.							
			[illegible] tons.							

PD-86 A
(7-8-42)

(Name and location of plant) (Department or operating unit)

SECTION E, *Part 2.*—This section must be filled out if more than one class of products is reported in Section B. See Instructions, Section A. Report all ma which entries appear in columns (5) or (8) of Section E, Part 1. Report materials in the units specified in column (3) of Section E, Part 1. Dollar v not to be reported in this section.

Item No.	Description of Materials (1)	Repeat Item No. (2)	Unit of Measure (3)	Consumption and Requirements for Separate Classes of Products — Products: Describe each class separately — A: Put into production April-June 1942 (4)	To be put into production October-December 1942 (5)	DO NOT USE (6)	B: Put into production April-June 1942 (7)	To be put into production October-December 1942 (8)	DO NOT USE (9)

for detailed types came from the same inventory bin. The problem may be explained more fully in this way. Suppose the Ajax Corporation reported in Section B, as follows:

Products	Dollar Shipments
Jigs	$10,000
Dies	10,000
Other machine tool accessories and parts	10,000
Electric conduit metal clamps	2,000

Each of the above items would be given a minor code. The first three, being items of the same genus, would carry one secondary code, and the fourth item another secondary code. Now, suppose Ajax either did not provide a Section E, Part 2 for any of these items, or combined the fourth item with one of the first three. Either of these alternatives would have raised a serious coding problem. The following details should have been reported for clean-cut tabulation: (a) use and requirements for dies, jigs, and other machine tool accessories and parts, and (b) use and requirements for electric conduit metal clamps.

In a situation such as this, several alternatives were open. The metal data for all products, including that for electric conduit clamps, could have been thrown into a product group the composition of which was solely devoted to information about machine tool accessories, since it was clear that the major products of the company fell into this category. For a small plant this would have had little importance. If the Ajax Corporation, however, had been larger than it is shown to be, a decision to do this would have seriously affected the accuracy of the over-all tabulations. It would have impaired the accuracy not only of the product group designed for machine tool accessories (which would have been loaded with other irrelevant data) but also of the product group concerned with electric conduit metal clamps (which would not have reflected the true volume of materials used and required for this product).

In view of this situation, of course, every effort had to be bent toward segregating information for both groups. This sort of problem was solved on the basis of information submitted by the plant on preceding questionnaires; from the knowledge of the Bureau of Census technical experts; from WPB Division personnel having cognizance over the plant; or from direct correspondence with the company. Frequently,

the schedule itself furnished information by which good estimates of the true situation were calculated.

Finally, a major code was given to each schedule to indicate the major product class of the plant or principal product as determined by dollar shipments. This code was necessary to tabulate receipts, use and inventory information for which no secondary product group tabulation was possible. The form did not provide for such a distinction of inventories. This decision in formulating the questionnaire, of course, was dictated by industrial practices and the tremendous difficulties which efforts to do otherwise would have imposed upon industry.

Other codes were placed upon each PD-25A schedule, such as a geographic code to permit tabulation by States and WPB Regions; a WPB Industry Division code, to show which section of the WPB would process the form and to provide it with a tabulation of concern to it alone; and various other codes designed for technical purposes, such as a completion code, a size code, etc. All of this coding procedure was dictated by the complexity of the final tabulations required.

Review and editing of PD-25A constituted a large and important operation. The principal functions of editing were to review information reported, to detect errors and correct them, and to prepare the schedule for machine tabulation. The report was surveyed to make sure that it was in usable form, was accurate and consistent with instructions, and would not disturb the over-all accuracy of the final tabulations. It is clear that tabulation of unedited or unreviewed schedules would have resulted in completely meaningless tabulations.

There were many instances where data reported on the PD-25A needed adjustment, but only a few of the less technical points need be treated here. Shipments were sometimes reported only in units rather than units and dollars; sometimes shipments for a future quarter were reported as smaller than those for a past quarter yet anticipated metal requirements for the advanced quarter were substantially beyond consumption in the past quarter; shipments reported on schedules of subsidiaries sometimes covered the entire company of which it was a part, while metal figures related only to the subsidiary; and so on.

Greater problems were encountered in connection with the metal figures. Sometimes manufacturers would report tons where pounds were required, and vice versa; frequently forgings and castings were reported in units, pipe in feet, and wire in rolls. Clerical errors on the part of plant personnel resulted in omission of important requirements and placing information on the wrong line or in the wrong column; and often odd material sizes and shapes were written in the preprinted stub.

Other typical and more difficult errors to correct were as follows: subcontractors, contrary to instructions, often reported material that they did not own but processed for a prime contractor; in some instances a manufacturer making basic metal shapes and forms specified on the preprinted stub, forgings for example, reported metal consumption and requirements for his finished basic product; manufacturers often reported one thing on PD-25A and in attached correspondence described conditions which necessitated eliminating the figures; and many times material reported in Section E did not appear in Section E, Part 2, when production clearly indicated the material had to be used to make the product specified. Obviously these errors had to be corrected for tabulation machines added integers only and produced figures no better than those which went into them.

In all tabulations it is important to know the exact part of the universe which the questionnaire is presumed to cover. Two problems in this connection presented themselves in tabulating PD-25A applications. First, it was necessary under the operation of the Production Requirements Plan to consider total requirements for critical materials and to compare them with total estimated supply. Since the PD-25A applications themselves were not used as application forms in certain non-PRP or "exempt" areas the WPB had to make arrangements to acquire "master PD-25A" schedules covering material use, requirements, and inventory for consumers in these areas. This was a WPB rather than a Bureau of the Census Problem. Second, in the absence of a long experience with the PD-25A and its predecessor PD-275, and because of changing war production and industry experience in filing this type of report, it was exceedingly difficult for the Bureau of the Census to know when sufficient reports were received to assure reasonably complete coverage in particular industries or for all industries. Tight tabulation time schedules necessitated stopping the tabulation procedure somewhere so that schedules could be added and presented to the WPB for its administration. But tabulations incomplete in totality or in important particulars could not be adapted readily to administrative needs.

Obviously one of the first important steps in assuring completeness was effective follow-up procedures for stimulating quick industry response. But beyond that the Bureau of the Census and the WPB had to develop measures of completeness for the entire tabulation and important product groups thereof. The most important and practical measure was a list of important plants that had to be included in the tabulations to assure completeness. The high degree of material con-

centration among a relatively few large plants facilitated the determination of an estimate of the extent of coverage of the total tabulation merely by checking the important plants included. Within narrow product groups this technique, although applicable, was not so practical. And no matter how carefully incoming schedules were matched against lists of important consuming plants, estimates of coverage in the aggregate or for particular products were subject to a wide margin of error and only the preliminary tabulations could reveal accurately the extent of completeness. This problem, due essentially to lack of historical series for this type of data, created an unnecessary risk in both tabulation and WPB operations.

Following the above processes, schedules were forwarded to the machine tabulation unit where punch cards were prepared. One card was punched for each line of Section B, one card for each item in Section E, Part 1, and one card for each item in Section E, Part 2. In addition, there was punched on each card the major product code, the secondary product code, the minor product code, the WPB Division to which the schedule was assigned, a size code, a completeness code, and a serial number for the schedule. Altogether approximately 500,000 cards were punched for the fourth quarter, 1942, PD-25A tabulations. In the next succeeding quarter the volume of cards almost reached 800,000. Preliminary tabulations were prepared from the cards and subjected to critical review. In this process, the figures were given additional careful scrutiny to detect errors appearing when the data for one homogenous group were listed together. Literally miles of tabulating machine listings were produced from these cards.

The foregoing account leaves little doubt that tabulating PD-25A schedules not only was a monumental undertaking but was made in the face of serious technical problems. The time schedule for tabulation frames the problems of tabulation *multum in parvo*. For the fourth quarter 1942 the time schedule was as follows: receipt of schedules, 25 October–10 November; first machine listing, 11 November; machine listing for the WPB Requirements Committee, 16 November–23 November. Here was what amounted to a census of the metal-working trades, completed in less than one month!

Despite the immense effort to adapt tabulation procedures to fit administrative methodology, the PD-25A tabulations and those of its predecessor PD-275, raised loud criticism in the WPB. Censure sprung from many factors, perhaps the most serious of which was the coding system used by the Bureau of the Census. The system of coding was necessarily rooted in a classification system which reflected existing

industrial organization. The development of the WPB Division structure, on the other hand, was in terms of product assignments which in many instances did not reflect the real organization of American industry.

These divergent factors created a fundamental conflict between the tabulated summaries and the WPB's ability to use them effectively. The net effect was a basic policy decision on the basis of tabulated summaries which could not be carried out fully by the WPB Divisions. The Census, for example, was obliged to tabulate insulated wire and cable into one category partly to obtain a clear-cut homogeneous group of data and partly to permit tabulation of individual plant schedules reporting production of insulated wire and cable, without reporting detailed data for all types and sizes. But in the WPB various types of insulated wire and cable were distributed and managed by different divisions with conflicts in the jurisdiction and responsibilities of each. The Copper Division was assigned various types of copper cable; the Communications Division was concerned with coaxial, lead-covered, telephone, and submarine cable; the Shipbuilding Division was interested in armored or degaussing cable; and the Building Materials Division processed requests for armored cable, electric cable, and insulated cable. Other products, such as electric motors, valves, and condensers were each under the cognizance of many WPB Divisions. Some important products such as storage batteries; turbines and water wheels; bolts, nuts, washers and rivets; and miscellaneous stamped and pressed metal products, were completely unassigned to any WPB Division. No juggling of tabulations could adjust to this illogical assignment of products within the WPB itself.

There were two further problems. The WPB Divisions, partly to fit tabulations into their organization and partly to fit them into their preconceived notions of the type of data summaries needed to provide efficient administration, felt that tabulated summaries should be made in much greater detail than was possible with the 230 product groups used in tabulation. A simple illustration of the problem relates to one product group established for "plumbing fixture fittings and trim." Into this group the Census tabulated products including victory trim which contained a small amount of copper alloy; brass trim which was 100 per cent copper alloy; plumbing specialities which might have been either; pipe hangers, supports, and rests, which were largely carbon steel; and other fixtures of varying types of metal content. The WPB problem centered about the policy determination of allocations of this group and Division and industry compliance with basic policy. With

such a farrago it was difficult for the Requirements Committee to know how much carbon steel included in the allocations for this group would be consumed in victory trim as against pipe hangers. Likewise the responsible Division had no method for determining, without special inquiry, what proportion of the metals allocated to this product code would be in victory trim (which was available to the public), or brass trim (which was available only to the military services). The results led to WPB pressures for finer and finer product detail. The Census problem in accomplishing finer product detail, of course, centered about limitations inherent in manufacturer reports. In the face of limitations imposed by the reports themselves the effort to produce more detailed summaries intensified the entire tabulation problem and subjected it more and more to error. In later quarters the product groups were refined but never to the complete satisfaction of the WPB Divisions.

A second type of problem pertained to the Census policy of attempting to make product group data homogeneous. But this policy, particularly in the first quarter of 1943 tabulation, met head on with increasing difficulties in the WPB in relating the chain of production to final products. The tabulating agent was asked, therefore, to code products into their end use categories whenever possible. This was done for the first time in the first quarter of 1943 PD-25A tabulation. Thus, for example, brass valves for ships, where identifiable as such, were set up in a separate product group under ships; valves for track-laying tractors were tabulated with track-laying tractors; low pressure hydraulic valves for aircraft went into aircraft; boiler feed regulator valves for locomotives were tabulated in locomotives, and so on. Once again, the principal limitation was in the plant reports themselves. Since a complete end-use identification was not possible for all components the resulting final summaries were rather far from homogeneous and did not completely reflect incorporation of component product data into product groups established for their respective end items.

All these difficulties were accentuated because of simple errors in the schedules themselves. Since the coding experts in the Bureau of the Census were few in number and since time pressures were so great, much of the coding had to be performed by clerks. Although errors were for the most part relatively minor, they assumed tremendous proportions in WPB debate. Natural unfamiliarity with a wide variety of trade names, many of which were introduced with war production, led to miscoding. Some homographic words were miscoded, such as concertina barbed wire under musical instruments and fire control instru-

ments under fire-fighting equipment. Such errors were magnified out of all proportion to their significance by WPB personnel and eventually led to rather widespread dissatisfaction with PD-25A and PRP.

Decentralized Tabulation of Operating Data under the PRP and the CMP

The PRP experience lent a deeper and stronger support to those WPB material distributive systems which followed it than most people are willing to admit. The CMP, for example, would not have been possible without the experience gained under the PRP. It is in the realm of tabulation that one of the supports for this conclusion is clearly revealed.

The PRP, as a mechanism for material distribution, brought to light many important "bugs," the spotting of which gradually filtered up through the various managerial levels of the WPB and on the way crystallized into rather clear-cut issues, which could be resolved only by top-side policy decisions. The more important questions related to the precise material shapes and forms to be controlled by a master distributive machine, the exact organization of summary data for policy decisions and administration, and the methods by which data were to be accumulated on both the requirements and authorization side of the management chain of action. These issues were resolved in the promulgation of the Controlled Materials Plan.

Conclusions of the planners and administrators of basic policy coincided in the fourth quarter of 1942 concerning the need for and advantages of decentralized tabulation of authorizations made by the WPB Divisions on PD-25A and other material allotment instruments. It was recognized that the WPB could no more administer its program for metal distribution without some accounting of action than a private business enterprise could carry on its day-to-day operations without some accounting for its cash withdrawals. In addition, it was recognized that decentralization of the authorization tabulations within each WPB Division offered many administrative advantages over centralized accounting control tabulations.

These concepts developed into a number of basic operating tenets. The following were of considerable importance: (1) those responsible for administration should also be made responsible for the figures which formed the basis of control action; (2) decentralized control records should be available for uses other than control accounting; (3) specialized problems relating to particular industries, plants, and products could be best ironed out on a local basis; (4) disagreements over methods of tabulation and accuracy of results disappeared under

decentrally posted but centrally directed tabulation; and (5) requirements data and authorization actions could be readily related through all levels of administration for each detailed segment of industry when data were tabulated decentrally. Undoubtedly, centralized tabulation had certain advantages over decentralized tabulation but administratively the latter technique was far superior to the former.

To test the statistical validity of these tenets both the Bureau of the Census and the WPB Divisions tabulated fourth quarter 1942 PD-25A returns. Substantial and ridiculously careless errors were detected in the WPB Division tabulations when the two series of data were matched. Nevertheless the experience was well worth the effort and firmly established both the utility and feasibility of WPB Division decentralized tabulation of data relating to the operation of a comprehensive metal distribution program such as the PRP.

In the meantime, plans were under way for the introduction of the Controlled Materials Plan in the second quarter of 1943. Tabulation techniques contemplated in the operation of the CMP fortunately were guided by the previous history of the PRP and other tabulation experiences. First, a firm decision was made concerning the metal shapes and forms to be included on the stub of the CMP application forms. Second, the WPB firmly grappled with the problem of product codes into which Class "B" products were to be incorporated. And they were fitted into the WPB's operations. This stilled much of the WPB criticism of Bureau of the Census product codes used in the tabulation of PD-25A. Third, the problem of proper editing of forms before tabulation was recognized and acknowledged. Finally, the WPB Divisions themselves were to make both the original requirements tabulations and the tabulation of actions taken on the basis of the WPB Requirements Committee decisions.

The problem of determining product groupings was divided into two parts, reflecting the combination of vertical and horizontal material distribution techniques upon which the Plan was based. On the vertical side, the Claimant Agencies determined their own groupings in terms of final products, or programs. The WPB problem with respect to the Class "B" list was comparable with that encountered under the PRP but of less formidable dimensions. Here classification had to be in terms of fabricated products entering into end products. To grapple with this problem the WPB established a Product Assignments Committee which attempted to reconcile the differences existing among the various Divisions, to establish firmly individual Division responsibility over an individual product, and to place responsibility for groups of

related products in one Division. Experience under the PRP went a long way toward defining properly the individual items to be included in a particular Class "B" code and in insuring that items calculated to be related were really homogeneous and reportable by industry in the terms specified. But despite this experience, there were some difficulties with the listing. There was some overlapping of responsibility among the Divisions, some product groups were too refined for good industry reporting and effective WPB processing, some product groups were too narrow and not representative of a homogeneous group of products, some products were completely overlooked, and some product groups were far too inclusive. Altogether, however, the difficulties raised in the original listing were not of substantial proportions and operations over a few quarters rather well ironed out the problems.

The classification scheme established in February 1943 for Class "B" products embraced 484 product groups as compared with 441 used under the PRP in the first quarter of 1943. This represented a substantial expansion because it will be remembered that the Class "B" listing was confined to presumably shelf items and did not include the industrial universe as did the PRP. A great volume of end products such as guns, tanks, ships, etc., were considered Class "A" products and taken care of in the programs established by Claimant Agencies. It is interesting to note that some PRP product groups were combined to form one CMP Class "B" group. Such instances, however, were few.

The history of product group classification up to this point was one of general expansion, modification, consolidation, and adjustment. This process was so widespread that any attempt to determine summary data covering the war period from the various instruments used by the WPB was greatly complicated by variations in product groupings, not to mention coverage, type of data obtained, and overlapping time periods. Thus, tabulation of combinations of reports was virtually impossible. The introduction of the CMP classification systems made data obtained from the series of comprehensive reports preceding it virtually useless in the operation of the CMP.

Preceding experience with authorization instruments demonstrated beyond doubt that to obtain good and accurate tabulations of requirements and to furnish the basis for accurate accounting it was necessary to subject incoming forms to some critical review. During the second quarter of 1943, the first quarter of the CMP operations, the Bureau of the Census tabulated CMP-4B requirements for the WPB. But in the third quarter of 1943, when the WPB Divisions first undertook that responsibility, the manual of instructions to the Divisions contained a

careful analysis of the types of adjustment necessary to provide uniformity in tabulation and program determinations. Recognition of and planning for inaccurate reports spelled the difference between usable and unusable WPB summaries.

Fundamentally, the problem of the WPB Division tabulation of controlled material requirements and authorizations against those requirements was simple. It was merely an additive function. The basic reporting requirements for specific metals and metals shapes and forms were well established; the problem of specifying end-use on the forms had been resolved by means of the extension of program symbols; the product classifications into which the data were to be summarized and assignments of the product groups to specific WPB Divisions had been accomplished with success; and finally, industry had ample opportunity to adjust its records to provide the sort of data needed on various application instruments. There was little else to do but add.

The WPB Industry Divisions were not prepared to tabulate material requirements on CMP-4B applications submitted for the second quarter of 1943, and the Bureau of the Census was asked to make summaries for the WPB. To facilitate operations the WPB received the applications, coded and edited them, and transmitted them to the Bureau of the Census for aggregation. Because of the preceding experience with the PRP, however, the WPB Divisions were quite able to maintain authorization accounts for the second quarter of 1943, and all subsequent quarters. In the third quarter of 1943, both the WPB Industry Divisions and the Bureau of the Census tabulated incoming applications, the dual tabulation constituting insurance against WPB Industry Division failure. As it turned out, a series of audits among the WPB Industry Divisions showed that tabulation was proceeding accurately and satisfactorily. As a result the Bureau of the Census tabulation was not completed. In the fourth quarter of 1943, WPB Industry Divisions assumed all CMP requirements and authorization tabulations with which they were concerned, except those pertaining to mill reports of shipments. These were tabulated by the Bureau of the Census.

Under the operation of the Controlled Materials Plan during the fourth quarter of 1943, which procedure continued throughout the war period with but minor modifications, the WPB Industry Divisions were responsible for a number of tabulations. The most important related to a summarization of material requirements reported on incoming applications, and accounts showing authorization made against them. The Industry Divisions tabulated data on incoming CMP-4B

FORM GA-298
FORMERLY IC-37
(7-28-43)

UNITED STATES OF AMERICA
WAR PRODUCTION BOARD

SUMMARY OF CMP-4B APPLICATIONS FOR ALLOTMENT OF
CONTROLLED MATERIALS FOR 4TH QUARTER 1943

DIVISION			
PRODUCT GROUP			
CODE	TOTAL NUMBER OF REPORTS SUBMITTED		
REVIEWED AND APPROVED BY	EXTENSION	DESCRIPTION	
CMP OFFICER	DATE	REPORT AS OF 1943	

SECTION C - PURCHASE SCHEDULE

CONTROLLED MATERIALS	CMP CODE	UNIT OF MEASURE	TO BE PUT INTO PRODUCTION 4TH QTR. 1943	REQUIREMENTS 4TH QUARTER 1943
STEEL .				
CARBON STEEL *(Incl. wrought iron)*	2001-2061	SHORT TONS		
ALLOY STEEL	2501-2561	SHORT TONS		
COPPER AND COPPER-BASE ALLOYS				
CU-BASE ALLOY SHEET AND STRIP	3011	LBS. OF ALLOY		
CU-BASE ALLOY ROD, BAR AND WIRE	3021	LBS. OF ALLOY		
CU-BASE ALLOY TUBING AND PIPE	3041	LBS. OF ALLOY		
BRASS MILL COPPER PRODUCTS	3051, 3061, 3071	LBS.		
WIRE MILL COPPER PRODUCTS	3101	LBS.		
FOUNDRY PRODUCTS	3201, 3211	LBS.		
ALUMINUM				
ROD, BAR, WIRE AND CABLE	4021-4121, 4131	LBS.		
RIVETS	4122	LBS.		
CASTINGS	4202-4218	LBS.		
SHAPES, ROLLED OR EXTRUDED	4251, 4301, 4311	LBS.		
SHEET, STRIP, PLATE AND FOIL	4351, 4361, 4601	LBS.		
TUBING	4401, 4411	LBS.		

SECTION A - VALUE OF SHIPMENTS ANALYZED BY PREFERENCE RATINGS AND CLAIMANT AGENCIES

ITEM	CALENDAR QUARTER							
	SHIPMENTS APRIL - JUNE 1943		ESTIMATED SHIPMENTS JULY - SEPT. 1943		ESTIMATED SHIPMENTS OCT. - DEC. 1943		RATED ORDERS SCHED. FOR SHIPMENT OCT. - DEC. 1943	
	($000'S)	(%)	($000'S)	(%)	($000'S)	(%)	($000'S)	(%)
TOTAL		100.0		100.0		100.0		100.0
PREFERENCE RATING								
AAA-AA1								
AA2								
AA2-X								
AA3								
AA4								
AA5								
ALL OTHER								
CLAIMANT AGENCY								
ARMY (OW)								
NAVY (N)								
A.R.C.O. (C)								
MARITIME (M)								
O.E.W. (E)								
O.L.L.A. (L)								
CANADA (D)								
N.H.A. (H)								
O.D.T. (T)								
O.C.R. (S)								
O.P.D. (R)								
O.R.U. (U)								
F.A.W. (P)								
W.S.A. (A)								
W.R.C.								
6 PROJ.								
OTHER AND UNIDENTIFIED								

*CLAIMANT AGENCY NOT TABULATED SEPARATELY - MAY BE INCLUDED WITH OTHER AND UNIDENTIFIED.

2-4878

applications for which they were responsible and summarized them for presentation to the General Statistics Staff of the WPB on form GA-298, shown in the accompanying illustration. These summaries were then duplicated, aggregated, and presented to the WPB's Vice-Chairman for Operations.

Some relatively minor complications in this simple additive process grew out of time pressures, the need for including large late respondents, and complicated summaries required from the Industry Divisions. There was a general tendency to increase the number and complexity of the types of information summarized by the WPB Industry Divisions. Tabulations for the third quarter of 1944, for example, were presented on forms GA-689 (Requirements) and GA-689 (Explanation). The format of these two forms presents eloquent testimony to the increasing diversity and complexity of the tabular requirements. But as the CMP operations became smoother the need for such complicated presentations vanished and in 1945 the tendency towards simpler reports again prevailed. The third quarter of 1945 CMP-4B requirements tabulations (made in June 1945), for example, reverted back to form GA-289. Although these were the basic summary presentations the responsibility still existed for the Divisions to provide a variety of other compilations.

In the operation of the CMP, as in the PRP, time schedules were always tight. In the fourth quarter of 1943, completed tabulations of almost 33,000 CMP-4B applications were made 55 days following the final date for filing the applications. Subsequent quarterly operations shaved this time cycle considerably.

WPB Field Tabulations

The real impetus to decentralized tabulations in WPB Field Offices originated with the broad decentralization policy of the WPB in 1943. Transfer of more and more operating functions to field offices at that time would have left a gap, had it not been filled, in the flow of significant information to officials responsible for the establishment of these policies. With expansion of field office actions, therefore, it became essential that they be charged with the responsibility for maintaining uniform reports on actions, consistent with and related to comparable records maintained in the WPB Industry Divisions.

Experience with decentralized tabulation, both in the WPB and its field offices, was satisfactory enough to set it apart as being feasible and practicable. Indeed, it may be concluded that sound wartime administration is better achieved when tabulation of data closely

related to administration is made decentrally by those who are responsible for the use of the data in management. This conclusion has great importance for mobilization plans of the future which must utilize decentralized administrative methodology.

PART VI

LESSONS OF EXPERIENCE

The task of the War Production Board and its predecessor agencies can be stated simply. It was to manage the industrial economy of the United States in such a way as to produce the maximum war and war-supporting goods and services with the minimum dispatch.

This was predominantly an administrative job. To an extent unprecedented in our history the national economy was guided by government between 1942 and 1945. During the war the construction of every new industrial plant was authorized by government. Not a pound of steel or copper or aluminum could legally be fabricated and used without government approval. In all industrial production and distribution the decisions concerning what should be produced, who should produce it, and to whom it could be shipped were directed by government.

Facts and proficient factual collection methods were indispensable to the intelligent and efficient performance of governmental management. Actions taken in industry were recorded on millions, yes billions, of pieces of paper. It became necessary for the WPB to aggregate information from these documents to provide the means for making informed policy decisions, to provide the instruments for administering the thousands of actions intertwined in a single policy decision, and to develop records of the actions taken on the basis of policy decisions.

Looking back on the beginnings in 1940 and 1941, and even the later developments of 1942 and 1943, two facts stand out in connection with the emergency data-collection system. First, many of those responsible for industrial mobilization and continuous control of the economy did not have a clear concept of the vital place which systematized factual reporting occupied in the emergency management function of government. Second, methodological discipline required to systematize the collection and use of facts needed for emergency management was not developed and ready for use. The result was the parallel evolution of a recognition of the need for factual accumulation to support management and the development of suitable techniques to get the necessary facts. The process of meshing the two was slow and costly.

Two classes of lessons emerge for those who in the future must deal

with the problems of emergency organization and direction of the nation's economy. The first group relates to the role played by an industrial statistical reporting system in emergency management. The second group relates to the technical methodology, and the principles which should be used in its application, in developing and presenting the facts needed for formulating emergency policy and administering that policy.

Before presenting these lessons it is worthwhile to draw attention to the applicability of World War II reporting experience to emergency problems of the future. It is our contention that no matter what the circumstances may be in a future emergency, wise management is impossible without facts; that it is possible to develop a system for collecting and using facts which is adaptable to government management of industry in all potential emergencies; and that the experience of World War II establishes the outlines and basic character of that system.

The Role of an Industrial Reporting System in Emergency Management

In the desperate search for means with which to supply the materials of war, we learned that effective decisions could be made only on a basis of knowledge. In the development of early control policies and techniques of administration it was assumed that statements of policy, supplemented by a few administrative reports, would provide the framework upon which industry could and would automatically mobilize for war production. It soon became apparent that the pull of the regular customer always diverted a substantial part of industrial output from war production unless statements of control policy were based upon understandable, complete, fool-proof facts and administrative techniques. Mere statements of policy, we learned, would not provide the guns, planes and tanks when manufacturers who supplied the goods of war were at the same time engaged in making automobiles, refrigerators, cosmetics and other items which offered them profitable use of materials, facilities and labor. We also learned that control over industry, once it converted to war production, could not be achieved by broad policy statements unsupported by concrete reporting to and from government. No firm control was possible without a method for collecting and using facts as instruments of both policy and administration.

The experience of the WPB clearly reveals the character of the industrial factual collection system which emergency management requires. Stated simply, it is necessary to have a complete and detailed

knowledge of the nation's resources and the demands which war will make upon them. This knowledge must be available on a comprehensive scale, reveal the most minute details, be in common terms by means of which the entire range of facts can be related to one another and to aggregates, and cover specific time frequencies. It means developing a national resources-requirements budget. But the factual collection system must accomplish more than that. Not only must it furnish the information essential for policy decisions but it must also provide the means through which mandates for specific action are transmitted to industry in specific terms. At every stage of responsibility this flow of authority must provide a basis for precise and definite accountability and reports of administrative action. These are the fundamentals of an adequate industrial reporting system suitable for, and flexible enough to support, emergency management.

Since the underlying problems of industrial mobilization are largely those of translating available men, materials, and machines into the greatest possible quantity of goods required for military operations and war-supporting activities, there must be a complete and penetrating knowledge of all resources. Information about resources requires a reporting system which will provide in integrated detail the actual and potential production of mines, forests, farms, smelters, refineries, factories, power plants, railroads, public utilities, and other productive facilities. The reporting system must provide knowledge of the smallest as well as the largest economic unit. The system should cover facilities for producing various types of materials, the actual level of operations, unfilled orders, potential capacity, labor requirements, material and component requirements, and so on. The entire fabric of industrial resources and their use must be revealed clearly in totality and in detail. The details must tie together in a factual picture, given in common units of measure and in definite time periods.

Against this knowledge of resources must be placed a statement of the demand for them in such terms that the two may be related. The demand for resources must be built upon a detailed, realistic and balanced statement of military, essential civilian war-supporting, and export program requirements. This statement must be in terms of specific items needed directly by the armed forces and civilians; the materials, components and facilities required for their production; and the time periods in which both final products and the intermediate products needed for their production are scheduled. In short, total requirements must be known and classified into procurement progams and production details. This factual statement must be realistic, ac-

curate, and in common units of measure which can be related to resources.

To determine the extent to which certain types of war activities could be supported, the WPB found that it required knowledge of the demands of a specific activity under consideration and information on all of the related supporting and dependent activities. Thus, in appraising the size of an ordnance program for Oerlikon guns, it was necessary to determine not only the quantities of nickel-bearing steel that would be used in the shafts of the guns but also the quantities of nickel-bearing steel that were required for the bearings of the gun, the shells that would be fired from the guns, the axles of the railroad cars that would haul the finished guns and the material used in their production. To appraise the one requirement for nickel-bearing steel all other requirements had to be evaluated, from the nickel-bearing steel used in the finished guns to the nickel-bearing steel required to make replacement bearings for a machine which mixed malted milks at a canteen in a war plant.

This sort of informational need grew out of the problems and circumstances of war. When the economic machine was straining at optimum capacity, when shortages appeared in many areas, wise direction of the machine was possible only when government on the basis of facts could choose among several demands focused at one point in the production process. If the supply of nickel-bearing steel was less than the total demand for it how else could an intelligent decision be made to double the quantity of steel entering into Oerlikon gun parts and reduce the amount of steel used as replacement parts for malted milk machines? Effective decisions required a factual review of total supply of and all demands for the steel.

It should not be assumed that once the size of the expanded war program was projected and the economic system was approaching capacity operations, the need for this interrelated information became readily apparent. On the contrary, realization of the need for facts and the magnitude of the problem of getting such facts was gained only after long and costly experience.

In the early defense period we tried to maintain all peacetime activities and at the same time satisfy the demands of war. Only when we realized how perilous was our existence and how great was our need for weapons did we have the courage to stop the production of a few unessential items. Later, in mid-1942, we belatedly recognized the fact that actual war output was too small and that we could not in fact have both guns and butter. Attitudes were rapidly switched by the

winter of 1942 and the military services were demanding programs which, if attempted, would not only have precluded butter but would have so strained the economy as to almost stop production of guns too.

In all of these actions we were driven continuously between desire and reality. The cord that joined the two was at all times knowledge. But the precise knowledge required did not exist when most needed. It was unfortunate that in the beginning we could no more foresee our needs for information than we could foresee our needs for armament. We tried to satisfy everyone in the early defense period by a free apportionment of materials and facilities and by the avoidance of a large-scale reporting burden on industry.

Ultimately we learned that we had to have facts and developed the outlines of the industrial reporting system required for emergencies. We learned that the system should produce a complete budget of all vital resources and all demands to be made upon them. The data must be complete and precise, from the supplies of basic resources to the procurement programs for final military, export, and essential-civilian products. With the data at hand it must be possible to adjust and balance programs among themselves and in the aggregate and to relate them to resources in such a way that objectives established are feasible of attainment. The data must show the time sequences in which various quantities of limited raw materials are needed at all levels of fabrication from basic material processes to final products. Throughout, the data must tie horizontally and vertically, must be expressed in uniform terms and must be related to uniform periods of time. Beyond all this, the methodology of collecting these data must be such that accurate results can be acquired in a matter of days. The more serious the emergency the more urgent becomes the demand for this sort of integrated reporting.

The more integrated and comprehensive the reporting structure the wiser will be basic policy and the more effective will be the administration of that policy. With an integrated reporting system it is possible to determine whether specific policy objectives or slightly modified ones can be achieved. It is possible to translate policy decisions into actions to be taken by large numbers of people at different administrative echelons. It is possible to convey in concrete terms to every person required to act both the specific authority and responsibility which each has and the precise limits imposed on the exercise of authority. An integrated statistical system should provide the methods and procedures by means of which information essential for policy decisions can be obtained, by which mandates for specific sections can

be transmitted through administrative levels, and by which precise accountability for the administration of each policy can be achieved.

It must be realized that an integrated industrial reporting system is not the sole basis for efficient mobilization and industrial control by government in emergencies. Time and again the WPB found that its internal organization had to be geared to the methods of the reporting system if the data upon which the WPB depended were to be used to the greatest advantage. Time and again the WPB had to adjust basic control procedures to fit the realities of sound data-collection methods. This does not mean that statistical methodology dictated either organization or control policy. Rather it means, as the WPB discovered, that since all three had to be related as closely as possible to the fundamental methods and procedures of industrial operations, all three had to be coordinated for the most effective operation of each.

Finally, as the WPB also discovered, the best statistical system will fail to reach its potential if there is not a free flow of information among all levels of management and if management at any important level fails to admit the importance of accurate factual knowledge in making decisions and administering them. An emergency statistical reporting system, even though it has the characteristics noted here as being important, will be no better than the degree of coordination of other elements of management and the intelligence with which it is operated and with which its data are used.

Applying the Technical Methodology of a Reporting System

In striving to produce the statistical discipline which it came to recognize as the basis of its operations, the WPB met with every important data-collection problem imaginable. In the solution of these problems it experimented with almost every known technique. Selected technical lessons which the war period taught are the substance of the preceding articles. Since the discussion is a concentrated condensation of a vast experience, it is not desirable to attempt further summarization at this point. A better purpose can be served by attempting to set forth a few of the more important principles which experience has taught should be kept in mind when applying the highly technical methodology of industrial factual accumulation in emergencies.

Experience shows that it is both necessary and possible to create a coordinated emergency industrial reporting discipline. An emergency industrial control agency is in a real sense an operating holding company over the productive machine of the nation. Such an agency

must be in a position to collect accurately and quickly at one central point or points summaries and details of the millions of actions which are taken in the highly complex production machine. The WPB's experience shows that it is possible to draw the factual threads of industrial action together at a central point or points in such a manner that the data are well coordinated, comprehensive in character yet penetrating in detail, accurate, and quickly tabulable for prompt use. If methodology is well planned, carefully applied and used with intelligence it can be made flexible enough to provide data to meet all important problems. The WPB never created such a statistical discipline but its experience leaves no doubt about both the need for it and the ability of government and industry to create it.

The development of a coordinated, flexible and useful reporting system is impossible in an emergency without some centralized control of the issuance of questionnaires and the application of technical methodology. In time of crises every pressure on administrators is toward immediate action. The drive for facts is so great that any promising method is attempted. The results of unbridled issuance of questionnaires in an emergency cannot be less than an unjustified burden on industry and chaotic government administration. By centralizing control over questionnaire issuance a focal point is provided for gathering together all sorts of questions relating to questionnaire policy and methodology. An opportunity is provided for such questions to be resolved in the light of the best policy and methodology. If this function is managed with wisdom the questionnaire system and its techniques may develop into and operate as a balanced fact-gathering mechanism. If it is not, the best plans and intentions will not create a coordinated data-collection system.

To approach a coordinated statistical methodology the many technical principles composing it and inherent in its satisfactory maintenance and operation must be well-formulated and clearly set forth. A balanced reporting system represents a highly complex structure in which details are of great importance. Not only may injudicious instructions, or poorly-phrased questions leave management without facts for crucial decision, but inattention to minute household details relating to such matters as mailing lists, follow-up procedure, editing, coding, and machine tabulation may also invalidate an important survey. And, paradoxically, the more complicated in substance and comprehensive in coverage the questionnaire the more important become these details to valid results. Questionnaires of small magnitude may be hand-processed to avoid administrative problems in govern-

ment. Comprehensive questionnaires cannot be so pampered. Slight error or miscalculation in any technical detail involved in their life cycle may become so magnified that the final results may be either suspected or ignored by management. The existence of a set of carefully considered technical standards which may be uniformly applied if the will and machinery exist to do it is a primary requisite to statistical discipline.

One of the most important groups of statistical standards which must be created and applied with the greatest skill is that relating to technical nomenclature and statistical units of measure. Not only are definitions and classifications of products, materials, plants, etc., crucial in the good use of data but in their determination there are present many problems which are administratively complex. Units of measure are, of course, basic in the functioning of a reporting system. Their determination and application also present exceedingly difficult and complicated problems. Statistics are a method of communication. They are, in addition, instruments of policy and administration. They are useless to serve these basic functions unless they are founded on a clear, understandable and administratively usable language of definition, classification and units of measure.

The WPB found that it is not important nor always desirable to embrace the complete universe of a subject matter in order to produce meaningful data. Concentration in industry of metal consumption, component usage, manpower requirements, and other production elements, is such that facts necessary to exercise control can be acquired from much less than the totality of units in a given area. For almost every product and material not more than 15 per cent of the total possible plant respondents account for not less than 85 per cent of the total economic activity for that product or material. By limiting reporting to these large units, and by exercising control in smaller units by means of general regulations, the objectives of emergency industrial control can be attained without imposing serious record-keeping burdens on the bulk of the small enterprises and serious clerical problems on government.

Reporting methodology must be geared to industrial practices if prompt and accurate data are to be acquired. The relationship between the design of centralized control procedures and methods of industrial operation must be carefully considered when developing administrative techniques. Experience teaches that any control procedure which accommodates itself to industrial practices can be instituted with a minimum of operational friction. Every new control system must go

through a warm-up period to familiarize industry with the new routines and to permit necessary adjustments to them. To the extent that operational changes are held to a minimum, this initiation period is shortened. In addition, any control system which is based upon industry's existing record-keeping practices is bound to yield better and more complete results. The closer the reporting framework and the controls which it supports approximate industrial practices, the quicker and easier it is to attain the objectives established for controls.

Better results are obtained if reporting is on a *quid pro quo* basis. Despite the obvious load which some complicated questionnaires, such as allocation and scheduling forms, imposed on industry, particularly on those plants whose record-keeping systems were not designed to yield readily the required data, reports were received quicker and with more accuracy than most experienced statisticians thought possible. On the other hand, many comparatively simple questionnaires designed to acquire general information for broad policy analysis were often met in industry with intense and bitter opposition. Response was slow and replies incomplete. The basic reason for this paradox is simple: the former eventually resulted in a right for the respondent to do something; the latter produced no direct advantages to the respondent. The obvious lesson to be drawn from this experience is that the more closely statistical data requirements are tied to application-type reporting forms the better the data will be both with respect to accuracy and completeness and the quicker returns will be made by respondents. In addition, industrial criticism and objections to a questionnaire, other things being equal, tend to decrease with the importance of the privileges or rights bestowed through the reporting instrument.

A comparatively long period of time is required to produce accurate and complete data through a new form. On the basis of the WPB's experience, from three to six months are required from the first submission of a new complicated reporting instrument to the time when both industry and government function reasonably smoothly on the basis of it. With advance publicity this period can be reduced, provided the questionnaire is not too complicated nor has in it injudicious procedures or questions difficult or impossible for industry to answer.

It is possible to reduce the number of data requests if a case analysis is made of the internal use of the data derived from the questionnaire and adequate provision is made within the control organization for the free and prompt exchange of information. The basic categories of information required for emergency management are not large. If sound technical standards relating to such matters as units of measure, no-

menclature, time periods and other details are carefully applied, and if a well-planned statistical system is used to support intelligent and suitable control policy, there is no reason why a relatively few reports cannot be made to satisfy fully management's requirements for facts. The experience of the WPB adequately supports this conclusion.

The adaptation of accounting techniques to statistical methods can be made effectively both to support implementation of policy decisions and to provide the necessary factual data for review of past decisions and new determinations. A statistical-accounting technique can exist side by side with a statistical reporting system and can constitute an important part of the entire data-collection framework. Although the WPB's material accounting was not as accurate as financial accounting, it performed for production control much the same functions as financial accounting performs for industrial financial control. By introducing tolerances permitted by material accounting needs, accounting control was instituted and operated at a far less expenditure of time and energy than financial accounting. Statistical accounting is a practical and necessary part of any emergency data-collection structure used to support control policy and its administration.

An adequate reporting system cannot be devised overnight. In early 1940 a tremendous mass of information was available in Washington pertaining to the operation of the industrial system. The data, however, could not be coordinated nor organized to present a comprehensive statement of resources. Data on potential war demands were for all practical purposes non-existent. The concept of an integrated emergency reporting system was unborn. Developing the concept and nursing it to maturity was a difficult and time-consuming task.

Experience of the WPB shows that it never fully recovered from these early statistical deficiencies. Makeshift schemes and organizations became deep rooted. It was impossible to erase them completely. As the war progressed the value of satisfactory new control systems were jeopardized because of existing inept procedures originated at a time when a scramble for factual knowledge resulted in the introduction of any procedure which appeared capable of producing some results. The erection of an effective statistical reporting system was, therefore, doubly handicapped.

Our experience after 1938 now demonstrates that we were as ill-prepared to create and use the proper techniques for collecting information as we were to wage modern war and produce quickly the weapons needed for its prosecution. Although we had a larger peacetime statistical organization than any other country in the world, we were not

prepared to provide the facts for an all-out economic effort. The reporting system merely provided a foundation. It was comparable in this respect to the world's largest steel capacity and the world's largest industrial capacity. We had the resources; we did not have the end-products that we needed.

War, we found, required not merely general facts; it demanded that facts be obtained on a broader basis, on a wider variety of subject matter, and in a shorter period of time than had ever before been necessary. To do this job we built upon prewar experience and techniques. Frequently, the needs of the emergency required substantial changes in available methodology. In some cases, emergency developments moved so far beyond original techniques as to almost preclude recognition of the foundation upon which they had been built.

For years we had collected biennially a *Census of Manufactures* which covered every fabricating establishment in this country. Data were collected for each establishment showing its employment, the products which it made, its consumption of electrical energy and the dollar value which it added by manufacture. The census was taken in odd-numbered years and became available to the general public some two years after the end of the period which it covered. In addition to this periodic census, we also had a large variety of current reporting series in the industrial field. Some of these were developed by the Federal government. To a larger extent they were the product of trade associations and business groups which cooperated in the collection of certain data necessary to their current operations. Over the years the Census data had provided a series of benchmarks against which current reports could be measured and projected. These were valuable statistics. For the purpose for which they were designed and used, they were undoubtedly all that was necessary.

The methodology and techniques were also adequate for the purposes for which peacetime industrial data were collected. In peacetime, data were collected to permit government and business administrators to make broad general judgments. Seldom were they used to administer specific actions. Business and government operated on the basis of a multitude of individual choices not directly related to and administered in conformance with national policy objectives. In wartime, the direction of the economy was founded on a single policy objective made by government. And, what is perhaps more important, the implementation of that specific central judgment depended on directions issued by the central government. Under such circumstances statistics became the basis of action. The data therefore had to be accurate and up-to-the-

minute. As a result, we could no longer operate with the methods of peacetime. A methodology had to be found by means of which penetrating industrial data could be collected in the most accurate and comprehensive terms and made available to administrators very quickly.

Our wartime experience demonstrated that no matter how urgent the need or how willing the cooperation, an integrated reporting system suitable for emergency management cannot be improvised or built in a short period of time. The development of useful information is a product of knowledge, skill, experience, and time.

If we are to profit from this experience and avoid costly errors in a future emergency, we must now determine upon a course of action which will permit us to preserve the accumulated knowledge of the past and to use it in molding a better system for future periods of crisis. We must develop not only the blueprint of the reporting system needed in an emergency, and relate it closely to emergency control plans, but we must insure that the major features of the statistical system are incorporated as far as possible in the current statistical reporting structure of government. A relatively small expenditure now to provide for emergency statistical preparedness, and if need be continued for the next 40 or 50 years, would be minute compared with the costs incurred in the three years of confusion attending the recent war effort, or the early period of confusion which we can be sure will be part of any future emergency program in the absence of a continuing pre-emergency effort to prepare an integrated emergency industrial reporting system.

BOOK REVIEWS

Edited by
OSCAR KRISEN BUROS
Rutgers University

Practical Business Statistics, Second Edition. *Frederick E. Croxton* (Professor of Statistics, Columbia University, New York, N. Y.) and *Dudley J. Cowden* (Professor of Economic Statistics, University of North Carolina, Chapel Hill, N. C.). New York 11: Prentice-Hall, Inc. (70 Fifth Ave.), 1948. Pp. xix, 550. Text edition, $6.35; trade edition, $4.75.

REVIEW BY ALFRED CAHEN
Textile Economist, Business Information Division, Dun and Bradstreet Inc., 326 Broadway, New York 8, N. Y.

THIS up-to-date revision of the popular original 1934 edition offers in understandable form some of the newer techniques on sampling, quality control, and time series analysis. Particularly effective are the step-by-step clear-cut solutions of numerous problems selected from business and government statistics. It remains one of the best texts on business statistics.

Some specific contributions notably well presented include Chapter 3, "Statistical Tables," and Chapter 4, "Rates, Ratios, and Percentages." This latter chapter effectively points out the inter-dependence of statistics and accounting. Chapters 5, 6, and 7 deal thoroughly with graphic analysis.

Time series analysis is effectively presented in Chapters 11, 12, and 13. Some simplified techniques are particularly useful as preliminary adjustment of data for price changes, calendar variation, and population changes.

Chapter 14 on correlation is perhaps oversimplified. The illustration on relationship between hardness and tensile strength of 27 pieces of wrought aluminum alloy results in a coefficient of +.991. The authors indicate in the preface that Chapters 1 to 15 may be used for a one-semester course and Chapters 16 to 22 for the second semester. Multiple correlation is dealt with six chapters later in Chapter 20. A realistic question arises whether presentation of simple correlation alone (separate from multiple) may not overimpress the student. "A little knowledge is a dangerous thing."

Chapter 16 deals with the normal curve. Happily the authors refer only once to coin tossing and not at all to dice or cards. These three illustrations are generally a curse upon both teacher and students of business statistics as they depart so far from reality. One example used to illustrate the normal curve is life experience of wooden telephone poles. Few business series can be fitted by a normal curve. It was originated by astronomers and generally applied to physical and biological data. The family of curves of the Pear-

sonian or Gram-Charlier series cannot be studied in a first year course. Therefore a chapter on the normal curve alone probably contributes little that a business statistics student will apply later in practice. However, the normal curve is important in the study of sampling distributions.

Some current indexes are analyzed in Chapter 21, consisting of the principal governmental price and production indexes. This chapter is simply and interestingly written. Perhaps it could best follow immediately Chapter 15, "The Construction of Index Numbers." The allocation of only one-half page to the logistic and Gompertz curves hardly seems adequate.

The final chapter, "Budgeting and Forecasting," is effectively presented except that no mention is made of unfilled orders and inventories—both of which are rather frequently used in forecasting.

The book contains 14 appendices: including mathematical tables as values of t, F, χ^2, etc; useful reference tables on sums of the first four powers of the first 50 natural numbers; squares, square roots, and reciprocals; logarithms; and interesting side lights such as rounding numbers and flexible calendar of working days.

So much for the high spots of the book. With so many worthy contributions, one may ask what are the limitations? That concerns primarily the use of the word "practical" in *Practical Business Statistics.*

Practical is defined by *Webster's International Dictionary* as "of, pertaining to, or consisting or manifested in, practice or action."

The emphasis in this text is primarily on techniques as indicated by the large amount of space devoted to the following topics:

SUBJECT	CHAPTERS	PAGES
Time series analysis	3	86
Correlation	2	54
Reliability and tests of significance	2	49

In contrast the authors are parsimonious on problems which their students will most likely encounter later in jobs in practical business statistics. Some of the larger requirements of the operating statistician are in the following fields (Note the number of pages given by the authors.):

SUBJECT	PAGES
Sources of data	8
Detection of errors	5
Questionnaires	5
Sample design (not including chapters on tests of significance)	7
Collection of data by interview or mail	2
Editing	1
Tabulation methods	6
Graphic presentation (adequately covered)	52
Report writing	—

These are realistic problems which both young and old practicing statisticians grapple with day by day.

Sources of data are the first step in tackling a business problem. The

authors do provide in Appendix A a five page list of general sources on production, price, wage and similar data. A concrete problem showing a student how to find the pertinent information on one industry such as wood furniture might be more constructive.

Detection of errors is probably the most important single problem which the practicing statistician encounters every day of his working life. The various checks and counter checks to help detect errors are scarcely mentioned in this text.

The selection of the proper tabulation method—hand, machine, or semi-mechanical—is highly important to the practicing statistician both in time and cost. Proper tabulation method sometimes saves several thousands of dollars on a single project. The six pages devoted to this subject by the authors oversimplify the problem by stating mechanical tabulation is desirable when a large number of schedules is being tabulated with many entries for numerous tables. This resembles the sales talks of the representatives of the International Business Machines Corporation and the Remington Rand Company. The determining factor in the use of mechanical tabulation is not the size of the project. When information needs to be both multiple sorted and repeatedly readded, then mechanical tabulation offers operating economics.

However, if it is largely a problem of multiple sorting without summations, the Keysort or EZ Sort systems can readily contribute less costly operation than mechanical tabulation for the same results. The reviewer saw an operation of Keysort at the Mutual Life Insurance covering 1,250,000 accounts—certainly a large scale operation. The authors do not mention the EZ Sort system which has a patented advantage of five holes to the inch compared with four holes for the older Keysort system.

In tabulations, regardless of size, where summation is the primary problem with little or no sorting, then the peg board system and the accounting machines of National Cash Register Company, both offer economics compared with mechanical tabulation.

Appendix B describes several calculating and adding machines. Crelle's tables offer a very economical means of multiplication, but are not mentioned. Although three different calculating machines and one adding machine are pictured, no reference is made to the comptometer. Various tests have shown that adding a set of figures twice on the comptometer is both more speedy and accurate than adding once on the adding machine and then reading back the tape. Tape reading involves the human frailty of an individual's mind wandering instead of listening to figures. Hence errors are not detected.

The authors have contributed much technical perfection and pedagogical skill in the revision of this volume. Some additional time could have been profitably spent to find out how their former students are actually putting statistics to use in practical business.

Dr. Juan Kimmelman rather aptly phrased the problem in the June 1944

Journal of the American Statistical Association "In many cases, after having devoted several years to practical work, even in high positions, he will admit that 80 per cent of the knowledge which he obtained in the university represents dead weight for him, and that it is neither needed nor of any possible use.

The examples in this text are selected from actual operations. However, the sources of these illustrations are largely from governments, institutions, and quasi-public giant corporations, the stratosphere of American business. They are *technical rather than typical* problems encountered from day to day by the rank and file of business statisticians.

All of the foregoing comments can probably be boiled down to a matter of relative emphasis. Formula problems are an essential part of a first-year statistics course. However, problems of collection of data, detection of errors, tabulation, and report writing are never given sufficient space and emphasis in a first-year statistics course.

The reviewer recommends this volume as a handy guide that every statistician will like to have on his desk for reference. Altogether it is an excellent text on business statistics—practical or otherwise.

Mathematical Methods for Population Genetics. *Gunnar Dahlberg* (Head of the State Institute of Human Genetics, Uppsala, Sweden). New York 3: Interscience Publishers, Inc. (215 Fourth Ave.), 1948. Pp. vii, 182. $4.50.

REVIEW BY HOWARD LEVENE

Instructor in Mathematical Statistics and Biometrics
Columbia University, New York, N. Y.

OUR present knowledge of genetic phenomena in human populations is still in a comparatively rudimentary state, and any advances are likely to prove of great benefit to the human race. The problems involved are: (*a*) to learn by investigation what sort of assumptions may reasonably be made about the processes which occur; (*b*) to formulate precise mathematical models embodying these assumptions and deduce their consequences; (*c*) to estimate the parameters involved. The Scandinavian countries, with relatively homogeneous populations, excellent vital statistics, compulsory reporting of hereditary defects, and governmental interest in the problem, offer a particularly good opportunity for investigation of problems (*a*) and (*c*). The author, who is Head of the State Institute of Human Genetics in Sweden, has played an important part in investigating all three problems. He has written a work with many excellences, combined with some defects.

This book is a translation, with minor changes, of the author's "Mathematische Erblichkeitsanalyse von Populationen," *Acta Medica Scandinavica*, Suppl. 148, 1943. The author has deliberately omitted all discussion of the work by the Fisher school and others on estimation of mode of inheritance, linkage relationships and gene frequencies for particular characters; and has made little use of the investigations of Wright, Fisher, Haldane and others

on evolution in natural populations. Within the author's chosen field, the book suffers from the defects to be expected of a book written on an essentially mathematical subject by a non-mathematician. Most of the proofs involve verbal reasoning and are consequently hard to follow and check. Furthermore, the precise mathematical model involved is not always clearly evident. In particular, the reviewer has felt for some time that geneticists use the term random breeding to cover a number of different processes that may lead to different results in finite populations, and the author's discussion seems to have this same fault. The author also fails to distinguish between the actual values of gene frequencies and their expected values, although at other times he makes effective use of the concept of random fluctuations in small "isolates." Furthermore, the modern theory of statistical inference is virtually ignored. In the section "Isolates and Racial Differences" a type of discriminant analysis is introduced which is in general inefficient in comparison with that of Fisher.

The above criticism should not be taken to imply that this is not a valuable book. For biologists and men of affairs it discusses clearly the problems of heredity in human populations, and shows the importance of further research along these lines. The final chapter in particular should serve as a valuable antidote to the more rabid eugenists and racists. (The author states that the first edition was intended in part for the post war edification of the German people.) The wealth of tables and graphs interpreting the various formulas should be particularly welcome to the less mathematical reader, and the conclusions drawn from them are likely to be at least qualitatively correct.

For the mathematical statistician and the biometrician, this book will be primarily valuable as an introduction to a new discipline of great practical importance which furnishes some problems of no mean mathematical difficulty. The reinvestigation and extension of the author's results, using the theory of stochastic processes and modern methods of statistical inference, provides a wealth of problems at many different levels of difficulty. All things considered, it is a book which will well reward reading by those with any curiosity toward this subject.

Principles of Biological Assay. *C. W. Emmens* (Head of the Department of Veterinary Physiology, University of Sydney, Sydney, Australia). Foreword by Percival Hartley. London W. C. 2: Chapman & Hall Ltd. (37 Essex Street), 1948. Pp. xv, 206. 21*s.*

REVIEW BY LILA F. KNUDSEN

Statistician, Food and Drug Administration

Washington 25, D. C.

UNTIL just recently, there has been no one comprehensive book on statistics in biological assay though the need for it has been great. There have been innumerable articles in scattered journals since most of the developments in this field have been in the last fifteen years, but nowhere have the

various statistical technics been gathered in one book. *Principles of Biological Assay* compares with an excellent recent book entitled *Probit Analysis* by Professor D. J. Finney of Oxford, which covers the field of biological assay with special emphasis on quantal (or percentage) responses and gives a rather thorough background in design and interpretation of such biological assays. Dr. Emmens' book attempts to cover a wider approach and includes quantitative as well as quantal responses on biological assays with one chapter devoted to the work of Finney and Wood on microbiological assays.

It seems that the title *Principles of Biological Assay* is misleading and should have been *Statistics in Biological Assay*, or maybe *Statistical Design and Analyses of Biological Assay*, since the entire book deals with the statistical design and analysis and not with the pharmacological aspects of assays as do other books in this field, such as J. H. Burn's *Biological Standardization* or Katherine Coward's *The Biological Standardization of the Vitamins.*

Dr. Emmens is primarily a pharmacologist who has found statistics extremely necessary in arriving at sound conclusions in experimental biology and the evaluation of strength and comparative effects of drugs. In 1939 he wrote a report on biological standards entitled "Variables Affecting the Estimation of Androgenic and Oestrogenic Activity" which was issued by the Medical Research Council of Great Britain. Dr. Emmens has come a long way, statistically speaking, since writing that report and his present book covers some of the newer important developments in statistical evaluation of biological assay such as Irwin's contribution for taking into consideration the error of the slope in calculating the fiducial limits of error, and Finney's probit plane analysis.

The book touches on the usual introductory statistical concepts; deals rather completely with dosage-response lines; describes various experimental designs (Latin squares, balanced incomplete blocks, etc.) that have been used in biological assay; gives formulas for calculating potency, standard errors, fiducial limits, combining results of several assays, chi squares, etc. However, emphasis seems to be on the manipulation of numbers rather than the assumptions made and the interpretations to be given the results. Even at that, the worker in the biological laboratory is given no inkling that in many cases it is possible to greatly simplify the formulas for routine calculations. The organization of the book seems a bit jumbled to the reviewer. The twenty chapters are each very short and each section does not seem to lead logically into the next section. The two chapters devoted to design of experiments are widely separated. Various tests of significance are scattered through the book. An entire chapter is devoted to a rather hazy explanation of polynomial coefficients and when factorial coefficients are introduced three chapters later, the author does not mention that they include the impressive polynomial coefficients. Student's t test occupies one paragraph and the only formula given is that for comparing a sample mean with the population value of the mean. Though the symbol t is used later in comparing two slopes and in calculating fiducial limits of the potency, nowhere in the book is an ex-

planation given of the assumptions underlying it and the way it should be interpreted.

The section on degrees of freedom seems unnecessarily complicated for absorption by a pharmacologist. The author apparently is trying to lead up to analysis of variance and factorial coefficients in two pages but simply confuses the reader.

Terminology can be very confusing. Dr. Emmens probably didn't realize that he gives three widely different concepts to C on pages 84, 86, and 105. It would have been possible to choose some other letters or symbols for two of the concepts. Another thing that may be confusing is the omission of the usual parentheses after the summation sign S in many equations.

It is rather surprising to see a book of this kind without lists of references for additional reading. It must have been an oversight that no mention is made of many of the numerous articles and statistical texts that would add to the reader's comprehension of the subject; although full credit is given in the text for examples cited and most formulas quoted. The method for evaluating biological assays given by E. B. Wilson and Jane Worcester in a series of papers published in the *Proceedings of the National Academy of Sciences* in the spring and summer of 1943 is not mentioned nor is Epstein and Churchman's "On the Statistics of Sensitivity Data" in the March 1944 issue of *Annals of Mathematical Statistics* nor Berkson's "logits."

It is consoling to find the statement on page 188 "The weighted mean log potency should have a variance $V\overline{M}$ given by

$$\text{(c)} \qquad \frac{1}{V\overline{M}} = S\,\frac{1}{VM}$$

which should not differ significantly from the alternative estimate

$$\text{(d)} \qquad VM = \frac{Su(M-\overline{M})^2}{Sw(r-1)}, \quad \text{where} \quad w = \frac{1}{VM} \text{ "}$$

and that "the best estimate of the potency available remains the weighted mean, which has a variance given by equation (d)." The implication is that even if the state of statistical quality control is known to exist the internal variance of an assay may give too small an error of the combined potency estimate. The left hand side of equation (d) contains an obvious typographical error. In the book it is written as VM instead of VM. The denominator of the right hand side of this same equation would be clearer if written $(r-1)Sw$.

On the whole, this book will be a welcome addition to the library of a pharmacological laboratory, particularly as a source book for formulas, and very helpful in promoting the use of statistics in biological assay.

Traffic Performance at Urban Street Intersections. *Bruce D. Greenshields, Donald Schapiro,* and *Elroy L. Ericksen.* Yale Bureau of Highway Traffic, Technical Report No. 1. New Haven, Conn.: Bureau of Highway Traffic, Yale University, 1947. Pp. xv, 152. Gratis.

REVIEW BY HENRY K. EVANS

Highway Specialist, Chamber of Commerce of the United States, Washington, D.C.

THIS 148-page technical treatise is a highly theoretical discussion of vehicular traffic movement characteristics (primarily at street intersections), comparing results of empirical field observations with mathematical calculations. The field studies employed a specialized photographic technique; the theoretical calculations revolve about the use of the laws of probability, especially the Poisson theory.

It is found that the field observations of vehicular spacings and groupings, (in moving traffic) agree very well with mathematical estimates of these characteristics. Therefore support is gained for the use of the mathematical approach in estimating traffic behavior, and the authors devote half of the book primarily to applying the laws of probability to different hypothetical traffic problems. The proof of the successful application of the mathematical method of calculating behavior patterns is well substantiated by the results shown in the book. However, as a useful tool to the practising traffic engineer or planner, the methods shown have no immediate practical value, principally because such precise knowledge of these behavior patterns, no matter how gained, bears negligible relationship to traffic engineering techniques commonly employed.

For example, various mathematical methods of calculating vehicle seconds delay at street intersections or drawbridges are presented. The traffic engineer is not as interested in knowing the precise amount of delay as he is concerned with knowing whether or not improvements in traffic control result in substantial reduction in over-all delay, which is easily observed by before and after field observations and simple arithmetic calculations. (The field observations would be necessary whether or not the mathematical refinements were applied.) Not enough is known about economic costs of delay to warrant any high degree of exactitude in determining vehicle seconds delay.

The highly theoretical aspect of the book is illustrated by the solution of a given problem of determining whether or not the frequency of obscuring vision (one driver's vision of a traffic island obscured by any vehicle within a 75-foot distance ahead) would be doubled by doubling the traffic volume. Use of Poisson's theory develops the estimate that 16 per cent of the drivers would experience obscured vision where traffic volume averaged 300 vehicles per hour, and that the per cent would be 29 per cent at double the volume. As a mental calisthenic this method bears some merit but has no practical use.

Another similar application of the laws of probability is made to a hypothetical traffic signal timing problem. Here the object is to find the shortest signal cycle length at an intersection which will pass the entering traffic volumes with failure to clear out the waiting vehicles only 5 per cent of the time. The method is useful primarily and only for the insight it gives into the traffic movement capacity of a roadway as limited by signal cycle length. Practically there are so many other factors bearing on selection of optimum signal cycle length that the theoretical method has little practical use.

Application of the Poisson theory to estimate the bunching of vehicles arriving at a parking lot or garage is brought out. This has a value to designers in making it possible to estimate the reservoir space required to care for vehicles arriving faster than they can be parked and is one of the most valuable applications shown of the mathematical approach to practical problems.

Whereas the techniques illustrated in this book have little application in solution of practical traffic problems, they do constitute valuable tools for research into generalized warrants for application of traffic control devices such as stop signs and traffic control signals. As yet the traffic engineering profession has not reached a verdict on what volumes of traffic or frequency of accidents justifies installation of stop signs at an intersection. The method shown of computing the percentage of unnecessary stops at a stop sign suggests that this knowledge might be used to determine that below certain levels of traffic volumes on two streets the percentage of unnecessary stops would be high enough to justify declaring that stop signs were unnecessary and a nuisance (provided other factors such as traffic accidents and view obstructions did not require such signs).

Although this book provides no ready answers or useful tools for the traffic engineer, it deserves to be acclaimed a milestone in the progress of traffic engineering research. It is the most scholarly and scientific investigation into the characteristics of traffic flow patterns yet published.

The primary contribution to the traffic engineering profession is the proved application of Poisson's probability formulae to analyses of traffic flow. Whereas it is doubtful that such application will prove of every day use for individual problems in setting signal cycle lengths, determining need for stop signs and other traffic facilities, or comparing relative efficiencies of different control devices, because of the many other non-mathematical factors affecting traffic flow controls and facilities; it is felt that the mathematical tools set forth in this report will be of inestimable assistance in performing basic research for the purpose of establishing general standards and warrants for different methods of control, sizes of traffic facilities, etc.

This method of traffic engineering mathematical analysis is akin to the first mathematical studies of nuclear physics. Whereas it has negligible immediate practical application, it is the entryway into a better understanding of the principles involved in traffic movement, which will inevitably result in practical application.

Theory of Probability, Second Edition. *Harold Jeffreys* (Plumian Professor of Astronomy, University of Cambridge, Cambridge, England). London E.C.4: Oxford University Press (Amen House, Warwick Square), 1948. Pp. viii, 411. 30s. (New York 11: Oxford University Press [114 Fifth Ave.]. $9.00.)

REVIEW BY HERBERT ROBBINS
Associate Professor of Mathematical Statistics
Institute of Statistics, University of North Carolina
Chapel Hill, N. C.

THE SECOND edition of this work on statistical inference differs from the first mainly in the addition of a theory of invariance aimed at establishing the logical consistency of the author's rules for setting up prior probabilities. The whole theory, based on Bayes's principle of inverse probability, lies entirely outside the modern development of statistical inference in the work of Neyman, E. S. Pearson, and Wald. Its relation to Fisher's work is more obscure; according to Jeffreys his results are in general agreement with Fisher's, although the underlying reasoning is not always the same.

Perhaps the most striking feature of Jeffreys' work is that his methods for testing hypotheses or estimating parameters take no account of the cost to the experimenter of accepting a false hypothesis or using an erroneous estimate. For example, a null hypothesis is to be rejected when the ratio of its posterior probability to that of the composite alternative is too small, the prior probabilities being determined by a set of conventions into which the cost of making errors does not enter. This procedure, which is based on an axiomatic concept of probability that has nothing to do with relative frequency, is not supported by any operational justification. Within Jeffreys' theory one cannot ask, let alone answer, whether the procedures are better or worse as operating rules than others which might be put forward as alternatives. This was indeed the situation in statistical inference thirty odd years ago.

Whatever one's views on the foundations of statistical inference, one cannot fail to profit by the mathematical ingenuity, keen physical intuition, and common sense which Jeffreys brings to bear on a wide variety of practical problems. Statistical theory will continue in large measure to find the inspiration and motivation for its real advances in the practice of the best statisticians. The appearance of the second edition of a book as provocative and rich in content as Jeffreys' *Theory of Probability* is a welcome event.

Rank Correlation. *Maurice G. Kendall* (Joint Assistant General Manager and Statistician, Chamber of Shipping of the United Kingdom, London E.C.3, England). London W.C.2: Charles Griffin & Co. Ltd. (42 Drury Lane), 1948. Pp. vii, 160. 18s.

REVIEW BY E. J. G. PITMAN
Professor of Mathematics, University of Tasmania
Hobart, Tasmania

IN THE words of the preface: "Until a few years ago rank correlation was a rather neglected branch of the theory of statistical relationship. In the practical field it was generally regarded, except perhaps by psychologists, as a makeshift for the correlation of measurable variables; and in the theoretical field it seemed to present no interesting or important problems. That situation has changed. Practical applications of ranking methods are not only being extended in psychology and education but are being made in other subjects such as industrial experimentation and economics. The theoretical properties of order-statistics have received much attention and are throwing important light on some difficult questions of statistical inference."

As the title indicates, this book deals only with the application of ranking methods to testing independence and to estimating degree of dependence of chance variables. The basic problem is, given n pairs of values of the chance variables X and Y, to decide whether X and Y are independent or not. By replacing the observed X values by their ranks, and the Y values by their ranks, we are able to test the hypothesis of the independence of X and Y without making any assumptions about the forms of the distributions. Moreover, we can do this when the values of one or both variables cannot be properly measured but can only be arranged in order of magnitude.

The book gives a clear and comprehensive account of the properties of Spearman's coefficient of rank correlation ρ, and of the coefficient τ, which depends essentially on the number of inversions in the ranks of one variable when the ranks of the other variable are arranged in natural order. Neither here nor elsewhere has the latter coefficient been given a name in English. The complications arising from ties are thoroughly treated.

The problem of m rankings, and partial rank correlation are dealt with, and also the relation of ρ and τ to the population correlation coefficient in samples from a normal bivariate population. The last two chapters discuss paired comparisons, as when an observer compares n objects two at a time without necessarily in the end arranging the n objects in a linear order. The tables necessary for making all the tests described in the book are provided.

Lack of precision in statement and in notation, and insufficient explanation sometimes make the author's argument difficult to follow. For example, on page 58 we have

$$
\begin{aligned}
E(c^2) &= E\{\sum (a_{ij}b_{ij})\}^2 \\
&= E\{\sum (a_{ij}^2 b_{ij}^2) + \sum{}' (a_{ij}b_{ij}a_{ik}b_{ik}) + \sum{}'' (a_{ij}b_{ij}a_{kl}b_{kl})\}
\end{aligned}
$$

where $\sum'$ denotes summation over values for which $j \neq k$ and $\sum''$ over values for which $i \neq k$ and $j \neq l$.

The summation in the first line is intended to include only terms in which the two first suffixes are the same or tied, and the two second suffixes are tied, i.e., terms like $a_{ij}b_{ji}$ are not to be included. It is natural to assume that the same convention applies to the first summation in the second line, and this is in accordance with the statement about $\sum''$. The author then goes on to show that the mean value of $\sum''$ is 0; but this requires $i \neq k$, $i \neq l$, and $j \neq k$, $j \neq l$. If that is so, the first summation must include terms in which the b^2 suffixes are the same as the a^2 suffixes, but in reverse order, so that terms like $a_{ij}^2 b_{ji}^2$ are to be included. (Note: $a_{ji} = -a_{ij}$, $b_{ji} = -b_{ij}$.)

Again, on page 61 it is stated that when summations extend over values of suffixes which exclude certain values (e.g., $i \neq k$) we can replace them by complete summations. This will puzzle the conscientious reader, as no reason is given. What is meant is that we can do this without affecting the dominant terms of their mean values. This is explained later, page 66, in the discussion of a more general case.

It is an excellent idea to publish a monograph on a selected portion of statistical theory, and it is to be hoped that other workers in statistics will follow this author's example.

Report on the Scheme for the Improvement of Agricultural Statistics. *V. G Panse.* Imperial Council of Agricultural Research. Delhi, India: Manager of Publications, 1946. Paper. 3*s*. 9*d*.

REVIEW BY S. LEE CRUMP

Iowa State College, Ames, Iowa

THE BROCHURE under review is arranged in eight chapters and four appendices. The main chapters discuss the following topics: statistics of area under agricultural crops, crop cutting experiments, forecasts and estimates of yield, sampling surveys for estimating yields. The remaining chapters are of a more general or incidental nature.

The total production of a crop in India is currently estimated from the following formula: Area $\times$Normal Yield$\times$Seasonal Factor. Dr. Panse has approached the problem of improving agricultural statistics with this formula as a take-off point. Each of the first two factors is discussed in considerable detail while the third is treated briefly.

The area under agricultural crops is assumed to be known with a high accuracy since an annual census is taken. The chief problems in this connection are those arising from mixed crops and from divided fields. Suggestions are made for reducing the error from these two sources. It seems to this reviewer that Dr. Panse may underestimate the importance of other errors which seem always to introduce incompleteness in censuses.

By far the greater part of the discussion is devoted to the problem of estimating yield. Dr. Panse has undertaken a thorough study of the methods of

estimating yield in current use. They all suffer from subjectiveness at almost every stage. The crop cutting experiments which should provide the best estimates of yield are not properly conducted, and are frequently ignored. Throughout the discussion the point is made that "normal yield" is an ill-defined concept and should be abandoned in favor of estimating the yield directly each year.

Finally, the results of a random sampling survey for estimating the yield of wheat are presented and discussed. As has always been the case, the random sampling survey method seems to offer the only economical scheme for obtaining reliable estimates.

The first three appendices give the details of the operation of the sample survey. In Appendix 4 two papers by V. G. Panse and R. J. Kalamkar on sample surveys are reprinted.

Dr. Panse should be commended for his approach to the problem. At each stage of the investigation he has attempted to suggest improvements in the estimates within the existing framework. That the final results indicate the overwhelming superiority of random sample surveys is not because Dr. Panse has arbitrarily rejected the existing scheme for making estimates.

Methods of Estimating Vital Statistics of Fish Populations. *William E. Ricker* (Professor of Zoology, Indiana University, Bloomington, Ind.). Indiana University Publications, Science Series No. 15. Bloomington, Indiana: Indiana University Bookstore, 1948. Pp. v, 101. Paper. $2.00.

REVIEW BY CHARLES M. MOTTLEY
Operations Analyst, Headquarters United States Air Force
Washington 25, D. C.

THE VITAL statistics of fish populations have undoubtedly been intriguing from the earliest times. The first fishermen must have wondered why the fish were abundant at one time and not the next. The surface of the water presents a barrier which man can seldom penetrate. The human census taker can draw representative samples, interview his subjects directly and gather, with relative ease, the information that he needs regarding the state of the population. The fishery biologist must resort to indirect census-taking methods if he wishes to determine the stock on hand or to measure fluctuations in abundance in relation to different environmental conditions or changes in fishing regulations.

The abundance of species that are exploited commercially is a matter of considerable economic importance. The accuracy of forecasts of relative abundance is of great concern to fishing interests. The California sardine fishing industry has just experienced an unpredicted and little understood period of scarcity which has caused considerable economic loss.

There is a growing list of publications on the subject of the vital statistics of fish populations and a considerable proportion of the funds for fishery re-

search is spent in this field. Dr. Ricker's work is one of the first attempts to bring together the body of modern knowledge on the subject. He has worked actively in this field himself. His own investigations have included original work on many species of freshwater and anadromous fishes. This first hand experience has been supplemented by a close study of the literature on many marine species, such as the Pacific halibut, the California sardine and many others which are referred to in his publication.

The subject of vital statistics is usually approached by fishery biologists through the data on the catch. After the introductory chapter, which defines the problems and presents certain important numerical relationships and terminology, Dr. Ricker devotes two chapters to the interpretation of catch curves. These curves show the relationship between age and frequency, and provide a means of gaining knowledge of recruitment and mortality. The age of fish can be determined by interpreting the growth markings on the scales or bony structures, or by noting the length frequency groups in the population.

A more direct method of deriving vital statistics is to mark some of the fish that are caught and return them to the water. The marked individuals can then be identified when they are recaptured. Tagging is frequently used, so that individuals can be recognized. The number of marked or tagged fish in the catch is used to calculate the rate of exploitation of the population and to determine the abundance. Dr. Ricker presents the different methods for computing these values and points out the limitations to marking experiments. He also lists the assumptions on which the validity of such population estimates is based. Dr. Ricker devotes six of his nine chapters to the fascinating problems of deriving vital statistics from the results of marking experiments.

He has included a useful "actuarial" appendix. It gives numerical values of the functions needed in this type of work, including the instantaneous mortality rate, the annual mortality rate and the annual survival rate.

Although Dr. Ricker's work will be of most value to fishery "actuaries," others will find it extremely useful as a source of ideas and methods. Military scientists have made use of Volterra's treatment of the problems of competition between species to develop combat models. Dr. Ricker's work may also prove to be useful in the apparently unrelated field of combat attrition.

The work is undoubtedly a first edition. It is hoped that it will be enlarged. In future editions some mention should be made of the approach suggested by Dr. D. B. DeLury and this reviewer, which utilizes the data on the catch per unit of effort to derive direct estimates of the total population under certain conditions.

Elementary Statistical Analysis. *S. S. Wilks* (Professor of Mathematical Statistics, Princeton University, Princeton, N. J.). Princeton, N. J.: Princeton University Press, 1949. Pp. xi, 284. Paper. $2.50. (London E.C.4: Oxford University Press [Amen House, Warwick Square]. 14s.)

REVIEWED BY T. A. BANCROFT

Associate Professor, Statistical Laboratory, Iowa State College Ames, Iowa

CRITERIA must be established for a critical evaluation of the plethora of new textbooks and new editions of old texts on elementary statistics appearing in the postwar period. The usual questions asked in textbook appraisals in general are: Is new material included, i.e., has the author taken cognizance of recent methodological advances and research applications? Again, has the material been arranged in a fresh and stimulating fashion without sacrificing soundness? Final judgment, of course, might ultimately rest on particular questions regarding clarity, simplicity, problems included, references listed, errors, type, general appearance, etc.

It is the reviewer's opinion, however, that, for texts on elementary statistics, a further question is in order, i.e.: What is the particular teaching program in statistics for which the text is designed? That is, was the text designed to be used as a beginning course in statistics in: (*a*) a single field of application such as economics, biology, agriculture, education, etc., or (*b*) in a mathematics department as another course in mathematics or a general service course for applied fields, or (*c*) in a department of statistics as a basic prerequisite course to advanced statistical courses or a general service course for applied fields? Presumably a text designed for a general service course for applied fields might be used either as a mathematics' or statistics' department offering provided mathematical prerequisites were the same. Also, a basic prerequisite course to advanced statistical courses might also use material adequate for a general service course. A companion question is: At what college-year level will the course be offered?

Professor Wilks has attempted with the material and arrangement of this new text to follow the recommendations of three distinguished committees on the teaching of statistics regarding the introduction of a basic elementary course available centrally to all students needing an understanding of statistical concepts and techniques common to all fields of application. Quoting from the preface, Professor Wilks says, "This book has been prepared for a one-semester basic course in elementary statistical analysis which, at Princeton, is the introductory course for all fields of statistical application, and is usually taken in the freshman year. It is especially designed for those who intend to go into the biological and social sciences. It presupposes one semester of elementary mathematical analysis covering topics such as those included in the first half of F. L. Griffin's *Introduction to Mathematical Analysis.*"

The text, then, is designed as a basic general service course, presumably to

be offered in a mathematics or statistics department at the latter part of the freshman year. It should be noted that the mathematical prerequisite implies an acquaintance with the elements of calculus which would apparently limit the availability of the text for freshman use to those colleges in which the elements of calculus are taught in the first semester or first two quarters of the freshman year. Within such a college the availability of the text would further be limited for the most part to freshmen whose curricula require the prerequisite mathematical analysis course. In some colleges this would eliminate altogether the students for which the text was primarily prepared, i.e., those in the biological and social sciences. In such cases a possible solution might be to follow still further the recommendations of the three committees mentioned earlier in setting up two basic elementary courses in statistics with different mathematical prerequisites.

A very good indication of the contents of the text may be obtained from the chapter headings: Introduction, Frequency Distribution, Sample Mean and Standard Deviation, Elementary Probability, Probability Distributions, The Binomial Distribution, The Poisson Distribution, The Normal Distribution, Elements of Sampling, Confidence Limits of Population Parameters, Statistical Significance Tests, Testing Randomness in Samples, and Analysis of Pairs of Measurements. No attempt has been made to include a discussion of the analysis of variance or more sophisticated problems in statistical inference since the material is designed for a one semester course only.

In alloting ten chapters to the elements of sampling statistics and only three to descriptive statistics, the reviewer believes that Professor Wilks has made a valiant attempt to provide pertinent material for a modern basic course to be centrally taught in elementary statistics. The text is remarkably free of typographical errors. No list of references is included. Although it is stated that the material is primarily for students in biology and the social sciences, many problems and examples are taken from engineering, industry, and the physical sciences. The reviewer believes that more attention should be given to experimental sampling and that a basic course in statistics should be conducted with benefit of a computing laboratory.

Professor Wilks' experiment in teaching such materials in a central freshman-level course should be observed closely. It represents an approach by a mathematical statistician to provide teaching material for this type course. Attempts to satisfy this need from another approach are being made at other institutions by applied or experimental statisticians. In the reviewer's opinion, the two approaches should supplement rather than compete with one another. Possibly ideal teaching materials for such a course will be the outcome of cooperative efforts of several statisticians with as many viewpoints.

LETTERS ABOUT BOOKS

Readers are invited to submit letters about statistical methodology books for publication in this forum. Concise, informative letters which supplement previously published reviews by pointing out specific strengths, weaknesses, errors, and errata in currently used books are wanted. Criticisms based on actual use of a book as a text are especially desired from statistics instructors. Other letters may consist of suggestions for the writing of books and reviews. Letters which contain adverse criticisms of JOURNAL *reviews will be submitted to the author of the review for any reply he may care to make. Contributors are requested to avoid personalities. The right to decide whether a letter merits publication is reserved. Letters should be sent to the review editor, Oscar K. Buros, Rutgers University, New Brunswick, N. J.*

EXPERIMENTAL DESIGNS IN SOCIOLOGICAL RESEARCH

IN THE literature of social science, two uses of the term "experiment" appear: first, to stand for trial and error attempts at the resolution of some problem of human relations (this is the popular usage); and second, to describe the method of precisely controlled observation used in physical science. It is to be expected that my use of the term, "experimental designs in sociological research," to mean observations of social relations under conditions of control introduced by matching, in my recent book, *Experimental Designs in Sociological Research* (New York: Harper & Bros., 1947), should be misunderstood and hence challenged in reviews by mathematical statisticians Kempthorne,[1] Ackoff,[2] Keyfitz,[3] and Hagood,[4] to mention a few. In fact, at least one sociologist, Hornell Hart,[5] has made substantially the same comment and suggested that the preferred term might be, "statistical comparisons with matched control groups."

The first use of the term "experiment," as a trial and error attempt to influence social relations by social action, I specifically exclude as the meaning of the nine studies summarized in my book (pp. 22–28, 29–33). My interest was to illustrate the crude beginnings of efforts to observe, under conditions of control by matching, what really happened to people when such trial and error experiments, taking the form of programs of social treatment or social reform, were used to influence them. It was my purpose to show that the systematic study of social action (trial and error "experiments"), is necessary if we are to appraise objectively the results often claimed for such programs.

The second meaning, I have also disclaimed (pp. 4–6, 26, 29, 32–33), although the term "experiment" was sometimes used as an abbreviation of the more cumbersome term, "experimental designs in sociological research," but when so used its reference was made clear in the context in which it occurred, even if the original author, whose work was summarized, used the term carelessly (Chap. 4).

These explanations and qualifications still leave us with the need of a term which may be used to describe studies of problems of social relations in which: (1) a group of persons who receive a program of treatment is compared with (2) a group excluded from this treatment; the situation is, (3) such comparisons take place in the natural community (not in the artificial situation of the laboratory or in the class room situation); and finally, (4) embrace an attempt to control by matching some of the factors in the situation other than, (a) the pattern

[1] Kempthorne, O. *J. Am. Stat. Assn.*, 43: 489–492, September 1948.

[2] Ackoff, Russell L. *Sci.*, 107: 509–510, May 14, 1948.

[3] Keyfitz, Nathan. *Am. J. Sociol.*, 54: 259–260, November 1948.

[4] Hagood, Margaret Jarman. *J. Am. Stat. Assn.*, 44: 312–313, June 1949.

[5] Hart, Hornell. *Social Forces*, 27: 96–98, October 1948.

of treatment factors, and (*b*) the pattern of response factors. Studies of this sort I have called experimental designs in sociological research. They are not, of course, "experiments" in Fisher's meaning of the term, and I have made no such claim.

The studies which I summarized were conducted in the complex community situation—hence the qualifying terms, "in sociological research." I suspect that the full import of these latter words can hardly be meaningful to any one who has not done empirical research in this field of complex social forces, although several reviewers[6] seem quite aware of their significance. I am still of the opinion that the designation, "experimental designs in sociological research" is a more meaningful descriptive title for these studies than Greenwood's[7] single term "experiment" (that promises too much) which he applied to several of these studies, or Hart's[8] "controlled comparisons" (which promises too little). No doubt usage will determine the acceptance and survival of these terms, as it does all other terms.

Inadequate definitions of the universes sampled is another point of criticism. As to this point, several of the investigators whose work I described did not, unfortunately, offer any such definitions. The chief explanation appears to be that the purpose of each study was to observe what happened to specific groups of subjects when treatment was or was not applied. The studies were exploratory and none was designed to provide a basis for generalization to a universe. Since, however, the criticism is well taken it may be worthwhile to note in passing that there are at least two types of universe concerned where social attitudes are measured. First, there is the universe of respondents in a defined area at a given time. Area sampling technique is the preferred procedure to be followed in all such cases. But two hypothetical universes are also involved and as yet little is known about the sampling techniques to be employed in such cases. One of these universes is that of all possible variations in the attitude of each person on a given issue over a period of time. This universe may be further subdivided into public attitudes and private attitudes, a problem of considerable practical as well as theoretical importance. Then there is the universe of all possible questions that may be formulated to elicit the true attitude of a person on a given issue. Stouffer[9] has recognized this problem by his statement, "Compared with the sampling of respondents, the developments in the sampling of items are still in a relatively primitive state." These points illustrate the great complexity of the concept of universe in sociological research.

The impossibility of randomization of the treated group of persons and the untreated group, presents the most serious problem, and is a well-deserved criticism if randomization has been neglected through ignorance, but randomization was not possible as fully explained in considerable detail (pp. 168–169). To state the problem briefly, no administrator of a private treatment program or of a legalized public reform effort would permit the recipients of treatment to be chosen at random. The local mores of the community or the public law that governs the program determines eligibility for treatment, whether it be by social case work, relief or public housing, by *need* for such treatment. In all such studies of programs we deal with a *de facto* situation which the observer cannot control. It would be an important contribution to research if some way could be found to get around this serious limitation. Furthermore, the consequences of such a limitation both as it precludes generalization from samples to a universe, and as it precludes control of unknown factors, is also discussed in detail (pp. 60, 78, 83, 89–90, 140–141, 166–169, 179–186) so that this point was by no means neglected. It may be also worth noting that because randomization was not possible

[6] See reviews by Hagood (*op. cit.*), R. F. Sletto (*Int. J. Opinion & Att. Res.*, 2: 412–413, Fall 1948), Donald E. Super (*J. Appl. Psychol.*, 33: 93–95, February 1949), T. G. Andrews (*Psychometrika*, 13: 281–283, December 1948), and Otis D. Duncan (*Rural Sociol.*, 13: 199, June 1948).

[7] Greenwood, Ernest. *Experimental Sociology.* New York: King's Crown Press, 1945. Pp. xv, 163.

[8] Hart, op. cit. p. 98.

[9] Stouffer, S. A. "Government and the Measurement of Opinion," p. 436. *Sci. Mo.*, 63: 435–440, December 1946.

no effort was made to use the powerful tool of analysis of variance. This restriction seemed to the author such an obvious logical consequence that no explanation of detailed reasons seemed necessary, although by implication at least, the reviews by mathematical statisticians suggest use of analysis of variance. Perhaps the best statement on this point is still that made by Margaret Hagood[10] in 1941, "Are the methods of analysis of variance (and covariance) applicable for observational situations where data are observed in the cross classifications in which they are found rather than in cross classifications in which they have been randomly placed in an experiment?"

Since the sociologist seems at this stage of development of his field research to be denied the advantages of analysis of variance, resort is made of necessity to the covergence of such evidence as he has: (1) the occurrence of small differences that are in the same direction of logical meaning (pp. 42–46, 49, 103–106); and (2) the persistence of such differences in more and more homogeneous samples after losses from mobility, death, refusals and inability to match have taken their toll. Patterns of difference which still persist after extreme cases have been lost, are regarded as evidence of real differences, and hence may be taken as evidence of the effects of treatment, subject always to the qualifications set by absence of control of unknown factors.

Questions naturally arise as to the validity of applying the conventional tests of statistical significance to differences between the statistics of samples that are non-random. Here again, I can but refer the reader to a somewhat detailed discussion of this point (pp. 176–186) wherein the qualifications and limitations are examined with the conclusion that only empirical results can be claimed, or as one reviewer puts it my solution is, "a pragmatic rather than a fully analytic solution."[11]

My concept of an *ex post facto* experimental design will have to rest on evidence yet to be gathered. Whether it will weather the tests of results in future research only time can tell.

Differences of opinion on the meaning of the concepts "null hypothesis" and "causation" appear in the literature of statistics and of logic. As to the null hypothesis my use follows that of Lindquist,[12] who appears to follow an earlier edition of Fisher's *Design of Experiments*, in which the concept was used to denote any exact hypothesis that one may be interested in disproving, and not merely the hypothesis that a certain parameter is zero, which later has come to be the conventional meaning among some groups of statisticians. Some other statisticians seem also to use the meaning I attach to the term.[13] For the sociologist, whose universes are complex and difficult to define, parameter values are seldom known; but there are real advantages in negative statement of relationships with the object of disproving them, because as I try to explain (pp. 70–73, 83–84, 93, 137–138, 167, 187) such formulations help us get rid of normative considerations. Criticisms of my use of the term seem to be based upon the contention that it departs from the current conventional usage in biological experiments and in biometrics. There are, however, examples of different usages of the same technical concept or term in other areas of research. For example, the term "ambivalence" has one meaning in chemistry and a different but accepted meaning in psychiatry; and the term "rationalization" has one meaning in systematic economics and a different meaning in abnormal psychology.

As to my use of the concept "causation," it will be noted in the first place that no proof of causation is claimed for any of the studies summarized (pp. 50, 74, *passim*), that replication of each study is recommended as the next step toward fuller understanding of the relationships (pp. 31, 57, 90, 120, 139, 176, 177, 185, 188, 189), and that my definition of causation (pp. 52–54) is consistent with the position of logicians as represented in the recent book of Herbert Feigl and Wilfrid Sellers, *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts, Inc., 1949).

F. STUART CHAPIN, *Professor of Sociology, Chairman of the Department, and Director of the School of Social Work, University of Minnesota, Minneapolis, Minn.*

[10] *Statistics for Sociologists*, p. 586. New York: Reynal and Hitchcock, Inc., 1941. Pp. viii, 934.

[11] Duncan, *op. cit.*

[12] Lindquist, E. F. *Statistical Analysis in Educational Research*, p. 15. Boston, Mass.: Houghton-Mifflin Co., 1940. Pp. xiii, 266.

[13] Hagood, *op. cit.*, p. 359; and Snedecor, George W. *Statistical Methods*, p. 54. Ames, Iowa: Collegiate Press, Inc., 1937. Pp. xiii, 341.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 44 | DECEMBER 1949 | NUMBER 248

ARTICLES

BOOK REVIEWS

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 248 DECEMBER 1949 *Volume 44*

STATISTICS OF THE KINSEY REPORT*

W. Allen Wallis
University of Chicago

In a preface that describes my own preconceptions and attitudes in approaching the Kinsey report, Dr. Alan Gregg of the Rockefeller Foundation, which "contributed a major portion of the cost of the program" (p. vii), observes that "the history of science is part of the history of the freedom to observe, to reflect, to experiment, to record, and to bear witness. It has been a perilous and a passionate history indeed, and not yet ended. . . . The findings of Dr. Alfred C. Kinsey and his associates at Indiana University deserve attention for their extent, their thoroughness, and their dispassionate objectivity. . . . Certainly no aspect of human biology in our current civilization stands in more need of scientific knowledge and courageous humility than that of sex. . . . These studies are sincere, objective, and determined explorations of a field manifestly important to education, medicine, government, and the integrity of human conduct generally. They have demanded from Dr. Kinsey and his colleagues very unusual tenacity of purpose, tolerance, analytical competence, social skills, and real courage. I hope that the reader will match the authors with an equal and appropriate measure of cool attention, courageous judgment, and scientific equanimity" (pp. v–vi).

* Alfred C. Kinsey, Wardell B. Pomeroy, and Clyde E. Martin, *Sexual Behavior in the Human Male*. W. B. Saunders Company, Philadelphia and London, 1948. Pp. xv+804. $6.50. The book is commonly referred to simply as "the Kinsey report," since Kinsey initiated the project in 1938 (p. 10), and did about four sevenths of all the interviewing that had been done up to the date of publication (p. 11). Similarly, it is common to write "Kinsey" where "Kinsey, Pomeroy, and Martin" or "the authors" would be correct. These abbreviated forms of reference are used throughout this paper.

This paper is a revision of one prepared by invitation of the 1948 Program Committee of the American Statistical Association and presented at Cleveland on 29 December 1948. Some of the material was presented to the Society for Social Research at the University of Chicago on 5 April 1948.

The notes have been collected at the end of the paper, since they should be read after reading the whole of the text. They are reasonably self-contained and can be read without reference to the text, but their numbers have been included in brackets in the text to indicate points that are expanded in the notes.

Sexual behavior is an important and interesting subject that has been considered intensively and extensively from nearly every viewpoint except the simple, factual one of what people actually do. Now, however, we have in the Kinsey report an intensive and extensive collection of facts about overt sexual behavior objectively defined. The report is, in short, a statistical study of sexual behavior. It presents figures on the varieties of sexual activity and the frequency with which each occurs. It reveals the variation of sexual behavior for given individuals and among different individuals, and it relates the variation to age, education, occupation, marital status, religion, and urbanization.

A statistical study cannot, of course, encompass the whole of what we think of as sexual behavior. A definition in terms of overt acts—Kinsey's definition is in terms of activities which ordinarily culminate in orgasm [1]—necessarily excludes aspects that some will describe as "the essential nature of sexual activity," the aspect chosen as essential depending on whether the chooser is an artist, biologist, criminologist, dramatist, psychologist, sociologist, theologist, poet, or philosopher, and perhaps depending also on his own sexual background. As Gregg says, "a great mountain may present aspects that are . . . so different that bitter disagreements can arise between those who have watched the mountain, truly and well, through all the seasons, but each from a different quarter" (p. v). Surely *one* of the significant aspects of human sexual behavior is the statistical one: how many do what, and how frequently? This is what Kinsey has undertaken to find out. That it is not the only thing worth knowing about sexual behavior is no ground for criticism; that it is an important thing to know about sexual behavior can hardly be denied.

There have been, to be sure, other statistical studies of sexual behavior. Kinsey describes nineteen (pp. 23–31), some of them excellent [2]. But no other study has been comparable with Kinsey's in the number of individuals included or in the amount of data about each individual. He has collected data from over twelve thousand persons [3], and has covered over five hundred items [4] for each. In contrast, 4600 is the largest number of persons, and 218 the largest number of items, covered in any of the other nineteen studies listed; and 2484 and 116 are fairer comparative figures, for the 4600 persons were studied superficially and the 218 items were coordinated with Kinsey's.

It is not by any means true, however, that Kinsey's conclusions are all based on statistical data. A great many assertions or implications about religious, ethical, sociological, psychological, and philosophical matters are scattered through the book—so many that I got a cumulat-

ing impression that the author is at heart a social reformer [5]. Most of his conclusions, explicit or implicit, about social and moral issues are based not so much on the statistical data "routinely secured in the interviews" as on "supplementary data" secured by other techniques. "These additional data have come from a considerable list of subjects with whom long-time social contacts have been maintained, in some cases for as long as seven and eight years. . . . While these supplementary records have contributed little to the statistical tabulations of data, they have provided a considerable portion of the detail . . . on the psychologic and social concomitants of sexual behavior, particularly in relation to factors which motivate and control the activities" (p. 74) [6]. It would appear, for example, that most of the material dealing with the attitudes of various social classes toward sexual techniques, their patterns of sexual behavior, and the social implications of class variations in sexual behavior (pp. 363–393) must be based on the supplementary data. In fact, much of the most interesting, and at the same time most controversial, material in the book appears to be based on the supplementary data. At least it clearly is *not* based on the statistical data [7]. One reason the many conclusions and interpretations based on the supplementary data have provoked controversy is that the data themselves are not presented. The passage on pp. 73–75, from which the quotation above was taken, seems to be the only account either of the methods of collection or of the data themselves.

The book contains three rather distinct types of material. There is methodological material, covering procedures for securing subjects, methods of obtaining information from subjects, the reliability of the data, and methods of statistical analysis. There is statistical data on the sexual behavior of certain groups of white American males. And there is "supplementary" interpretative material of a sociological or cultural anthropological kind. These three types of material are distinctly different in pertinence and in scientific quality. Only the second pertains primarily to the subject matter described in the book's title, and only the second is a scientific report, in the sense of an attempt to set forth systematically not only conclusions but the evidence on which the conclusions are based and from which other investigators can plan independent programs for verification or extension. The methodological material has some of the aspects of scientific reporting, but it contains in addition many unsubstantiated assertions [8]. The sociological material consists essentially of insights and opinions about sex in our society. Its character is somewhere between intelligent, technically trained, deeply interested, and thoughtful observation, like Barnard's

work on administrative behavior, and systematic but perspicacious anthropological field work, like Redfield's work on the culture of Central American villages [9].

The three types of material are neither differentiated nor integrated in the volume. The methodological material refers not to the "5300" males to whom the statistical results relate, but to all twelve thousand individuals so far studied. Conclusions based on the sociological interpretations or the supplementary data are frequently stated along with those based on the statistical data, and it is frequently difficult to judge what the basis is for a given conclusion. A clearer distinction would not only have added scientific stature to the report but would have eliminated, or at least focused properly, much of the criticism and some of the enthusiasm that has greeted it.

The book is divided into three parts: (I) History and Method, (II) Factors Affecting Sexual Outlet, and (III) Sources of Sexual Outlet. While this organization results in some unavoidable duplication between the second and third parts, it is on the whole worthwhile to be able to find all the material on, say, the relation of age to sexual outlet brought together in a single chapter of Part II. Similarly, it is an advantage to have all the material on, say, marital intercourse brought together in a single chapter of Part III.

When I first examined the volume, paying attention mostly to its fascinating substantive findings and scarcely at all to its methods, I was very favorably impressed indeed. When I diverted my attention to the general methods I began to note shortcomings; but I felt that these were technicalities—mere blemishes on the surface of the monument, which might modify some of the findings in detail but surely would not affect the broad conclusions. After all, many of Kinsey's figures would still be important and interesting even if we had to allow for an error factor as large as two or even three. But when I spent some time studying the statistical methods in detail, I realized that my confidence in the basic significance of the findings cannot be securely buttressed by factual material included in the volume. In fact, it now seems to me that the inadequacies in the statistics are such that it is impossible to say that the book has much value beyond its role in opening a broad and important field.

Instead of emphasizing directly the reasons for these misgivings about Kinsey's statistical techniques, let me bring them out indirectly by describing some of the improvements that I would like to see in later volumes. Kinsey plans several more reports, and there is every indication in this first one that he welcomes and even seeks criticism

that might help him overcome the numerous imperfections in his work, which he is the first to recognize; so we have here, for once, an instance where there is really a good case for "constructive" criticism. While it is not feasible to make specific detailed recommendations about statistical work unless one is in close touch with all the practical circumstances surrounding it, it is nevertheless possible to suggest tentatively directions in which improvement is possible.

It will be convenient to group my suggestions under three headings, relating to the collection, to the presentation, and to the interpretation to the data.

COLLECTION

I will not discuss the problem of determining the sexual behavior of a given subject—that is, the variables selected, the interviewing technique, and the list of questions [10]—but will confine myself to suggestions concerning the selection of individuals.

The question of sample size ought to be thoroughly reconsidered. Kinsey emphasizes that "the chief concern of the . . . study is an understanding of the sexual behavior of each segment of the population, and that it is only secondarily concerned with generalizations for the population as a whole" (p. 82). He defines his segments according to 12 criteria (sex, race, marital status, age, age of adolescence, education, occupation, occupation of parent, religion, religiousness, urbanization, and geographic residence), each having from 2 to 10 categories. This makes 384,912,000 segments if my arithmetic is correct or "nearly two billion" (p. 81) if Kinsey's is [11]; but as Kinsey points out most of these are fortunately non-existent or rare, and actually only 163 (p. 29)—an impressive enough figure—are covered by the data in the book. Over 40 pages is devoted to the problem of how large a sample is necessary in each segment, and the conclusion is "that a sample of 50 has proven adequate for establishing incidence data, that samples of 100 or 200 are fairly adequate for means and medians, and that samples of 300 or more are quite adequate for determining means and medians . . . but . . . smaller samples may still be taken as indicative of results that may be obtained from larger samples" (p. 683). Unfortunately, however, the discussion of necessary sample size is unadulterated nonsense [12], and represents a prodigious waste of effort.

A proper determination of sample size for Kinsey's material will not be simple, though it should be much easier now that he has collected twelve thousand histories than it would have been earlier. The desirable sample size depends on the requisite accuracy of the results, for one

thing. To specify accuracy, as Kinsey does (pp. 83–85, 736), simply by some arbitrary percentage of the true figure is ambiguous for proportions and may be impracticable for averages, so even the specification needs further consideration in later work. It may be that a double-sampling procedure would be appropriate; perhaps the sequential estimation procedure recently developed by Charles Stein would be of real value [13].

In considering sample size for future work, account must be taken of the fact that many of Kinsey's cases represent the same individuals, and are therefore not independent [10]. Another technical difficulty in determining sample size is that some of Kinsey's sampling is cluster sampling—that is, he selects certain "groups" such as sororities, hitch-hikers, or mental institutions, and interviews all members of the group—so the individual histories in his sample may not be independent.

I mention these difficulties not to imply that Kinsey should not treat each history as many observations or that he should not use cluster sampling (though I do feel that both techniques require further analysis [14]) but only to indicate that a proper determination of adequate sample sizes is not an entirely simple problem. A further complication is that multiple measurements are involved in each observation, but this is characteristic of sample-size problems.

Consider next the composition of the sample. Kinsey states that it is "valid to extend generalizations" from his samples to the "163 groups on which data are given" (p. 29). Now a *sine qua non* for generalizing from a sample to a population is randomness. But, as Kinsey points out, in a human survey it is impossible to define a population clearly and then produce a sample that can be guaranteed random (pp. 92–93). On the other hand, any batch of data that we do get is a random sample from some population; that is to say, indefinitely many repetitions of the procedure that produced the sample will produce a population. We have two handles to manipulate in generalizing from human samples: one is to define our population as closely as we can and then attempt to approximate randomness in our sampling procedure; the other is to analyze our actual sampling procedure as well as we can and then attempt to describe the actual population to which it relates [15].

In my judgment, Kinsey in his future work should devote a substantially larger part of his resources—which I realize are limited—to attempting to get a good grasp on one or both of these handles. On the basis of this first report, I believe that far more can be done. With respect to actually defining his populations and then sampling them at

random, the following remark is suggestive: "we have a network of connections that could put us into almost any group with which we wished to work, anywhere in the country" (p. 39). This network together wih the wide publicity that the book has received, may reduce the necessity of spending "days and weeks and even some years . . . in acquiring the first acquaintances in a community" (p. 39).

But, whatever success may be attained in this randomization surely will fall short of perfection. At least it falls short of perfection in such relatively simple matters as marketing surveys, income studies, and even presidential polls. So every step in the sampling process should be paralleled by two steps aimed at studying the sampling process. How many refusals are encountered? How does the refusal rate vary among segments? What are the determinable characteristics of refusers in the various segments? It is said that "the restrained histories have, on the whole, been the more difficult to get" (p. 103). What is the evidence? How much harder? Are there trends in the histories with respect to restraint or other characteristics? It is said that many of the subjects have cooperated in order "to obtain information about some item affecting their personal lives, their marriages, their families, friends, or social relations" (p. 37). Are records kept of these questions, and if so are the questions related to patterns of sexual behavior? It is said that "the greatly disturbed type of person who goes to psychiatric clinics has been relatively rare in our sample "(p. 37). What is the evidence? What similar information is there about the personality types included in the sample, and what more can be obtained? Rare relative to what?

I should judge that nearly half the total cost of analysis for a project such as this would be in checking on the population-sample relationship; but it is worth it, for the validity and usefulness of the research depends fully as much on this as on the soundness of the actual measurements.

A final remark with respect to data collection: In future work it is desirable to plan carefully in advance the methodological checks that are to be included, as for example retakes, comparison of spouses, comparison of interviewers, comparison of remote and immediate recall, comparison of earlier with later results by the same interviewer, analysis of intra-cluster correlation, etc. If in these comparisons the factors not directly involved were balanced out, the results would be substantially more meaningful than are those presented in the book. Furthermore, it is well to avoid comparisons in which, as in the investigation of the clusters (pp. 93–102) or the comparison of age groups (pp. 198, 200) one sample is compared with a second which

includes the first; instead, the first should be compared with the remainder. Application of the statistical principles of experimental design would make it possible to carry out the methodological checks both more effectively and more economically.

PRESENTATION

My strongest recommendation about future reports is that they should tell precisely what was done. My strongest complaint against the present volume is that when I study it in any detail I frequently cannot tell what information Kinsey's conclusions are really based on.

In the first place, I hope that the next volume (or at least a mimeographed supplement to it) will give an account of the actual questions used. There are admittedly difficulties in doing this, for the questions are numerous and "have never been standardized" (p. 51) because "the form of each question has varied for the various social levels and for the various types of persons with whom the study has dealt." For example, "sexual vernaculars must be used in interviewing lower level individuals" and "such vernaculars vary considerably among different groups" (p. 52). Nevertheless, the point which each question covers is said to be "strictly defined" (p. 51) and terms such as "petting" and "prostitute" are precisely explained to the subjects. The interpretation of nearly everything in a study like this depends upon the questions, and their absence makes it impossible for other workers independently to test or to supplement Kinsey's data.

In the second place, future reports should show the actual number of histories involved and how they are distributed with respect to all the controls—age, education, occupation, religion, etc. [15] As a matter of fact, I am not even sure how many white males are covered by the present volume [3]. Furthermore, every effort should be made to present the basic data, at least in a supplementary monograph. Such a presentation should show for each question and each segment the responses given. It is not too late to publish such a supplement to the present volume; this would require a table of 163 columns and perhaps 300 rows, and could probably be presented in less than a hundred pages.

Any information given on the basis of all histories so far collected, as in methodological tests, should also give separately the corresponding figures for the histories covered in the specific report. Obviously statements that are true of twelve thousand cases may not be true at all of a specific set of 5300.

In general, I found the tables apparently meaningful while reading

the book for its substantive content; but when I studied them I found many of them confusing and quite a few downright unintelligible [16]. This is intolerable in a statistical report, but it can be avoided easily in future volumes by competent technical editorial work.

The explanations of statistical techniques struck me as thoroughly unclear. The explanation of the Accumulative Incidence Curves, described as "the one new statistical tool which we have had to develop for this study" (p. 114) and its superiority over the ogive, would have left me bewildered had I not come across essentially the same idea clearly explained in the June 1946 *Journal of the American Statistical Association* in an article on "The Operating Life of B-29 Engines" by Oscar Altman and Charles Goor, who casually refer to the technique as a standard actuarial device. And Kinsey's formula for the median (p. 113) is just wrong [17]. Again, these things are intolerable in a statistical report; but they can surely be remedied in future work.

As for the so-called "U. S. Corrections" (pp. 105–109), I have been unable to understand them. The explanation sounds straightforward, but a few calculations I have made with them do not check with Kinsey's [18]. The U. S. Corrections are obviously intended as a set of weights for combining various of the 163 segments into averages applicable to broader groups, and of course some such set of weights is needed.

INTERPRETATION

Most of what I have to suggest under this heading has been covered by implication under the "collection" heading. In general, if a statistical investigation of this kind is well planned and the data properly collected the interpretation will pretty well take care of itself. So-called "high-powered," "refined," or "elaborate" statistical techniques are generally called for when the data are crude and inadequate—exactly the opposite, if I may be permitted an *obiter dictum,* of what crude and inadequate statisticians usually think.

Kinsey's data have very properly been subjected to a minimum of processing. His measurements of incidence, frequencies, and accumulative incidence are straightforward and sound, and no doubt his U. S. Corrections are too, if I only understood them. While he smooths most of his curves, he always shows the original data too [10]. All of this I hope to see continued in later volumes.

The only measure which is seriously mishandled is the range. This is a statistic which can be interpreted only if one takes into account the number of observations, makes stringent assumptions about the

normality or other mathematical characteristics of the population and about the independence of the observations, and has at hand an appropriate set of tables relating sample range to population variability. If the variability in several populations is constant, the range in samples will be larger the larger the sample. Kinsey falls into a fallacy when he shows (p. 234) the next-to-highest frequency of outlet found at various ages and interprets the marked decline with age to indicate that variability declines with age. It would not be surprising if it does in fact decline with age, but the table does not show it without further interpretation, for the number of cases declines from 3,012 in the youngest group to 58 in the oldest. Sample ranges should not be shown in future studies but some other measure of variability should be substituted, for example the standard deviation, the average deviation, or—probably more practicable—the distance between the first and ninth deciles.

It would be desirable in further work to distinguish between differences that are not statistically significant—in other words, that aren't there as far as the data show—and those that are statistically significant but of little consequence. In several instances the present book disregards differences which the data establish, but which if admitted to exist could be shown to be small enough to be negligible for many purposes. For example, different interviewers seem to get significantly different results (as is characteristic of many measuring devices), but the differences are apparently not great enough to affect the conclusions seriously (pp. 133–143).

Terman's criticisms of Kinsey's inerpretation of the data, and also his criticisms of other aspects of the work, all seem to me entirely sound [19]. Terman maintains convincingly that Kinsey has misinterpreted his data when, for example, he concludes that "the sexual patterns of the younger generation are . . . nearly identical with the sexual patterns of the older generation in regard to . . . many types of sexual activity" (p. 397), and when he asserts that sexual patterns are stable throughout life, assuming in childhood the pattern of the occupational group to which the individual will ultimately move (p. 419). But it is hard to make specific recommendations for avoiding this kind of misinterpretation, except to reiterate that a substantially better quality of statistical work is essential if future research is to have value.

After studying the volume I still feel that it is a pioneering and monumental work in an important field. And of the authors' virtues as listed by Dr. Gregg, I still admire their tenacity, their tolerance, their social skills, and their courage; but I have major reservations

about their analytical competence insofar as that means statistical competence. The work will take a place in the history of this subject, it seems to me, analogous to that occupied in the history of price studies by Thorold Rogers' seven volumes on the *History of Agriculture and Prices in England, 1259–1793*, of which the *Encyclopedia of the Social Sciences* says, "Even his severest critics express admiration of his scholarly labors in extracting prices from such sources as the bailiff's accounts of the property held by the colleges of Oxford and Cambridge and the great monastic corporations of the Middle Ages. It is the interpretation of these figures . . . which is open to question and correction." "From the point of view of modern statistical technique these studies leave much to be desired. They have been modified and supplemented by recent works, especially those of Beveridge, Usher, and Hamilton, so that it is now possible to trace with some assurance the general trend of prices in western Europe through the interesting period of the 'price revolution'" [20]. As a consequence of Kinsey's labors it will no doubt be possible at some future date to describe with some assurance the statistical pattern of sexual behavior.

NOTES

[1] The following passage comes as close to defining sexual activity as any I have found in the book: "The sexual activity of an individual may involve a variety of experiences, a portion of which may culminate in the event which is known as orgasm or sexual climax. There are six chief sources of sexual climax. There is self stimulation (masturbation), nocturnal dreaming to the point of climax, heterosexual petting to climax (without intercourse), true heterosexual intercourse, homosexual intercourse, and contact with animals of other species. There are still other possible sources of orgasm, but they are rare and never constitute a significant fraction of the outlet for any large segment of the population" (p. 157). "Outlet," as used in the report, seems to be equivalent to "orgasms" (pp. 193, 683). The six forms of behavior listed constitute the sexual behavior studied in the report; two measurements are presented for each type: (1) incidence, the proportion of individuals engaging in the activity, and (2) frequency, the number of orgasms per week from the activity.

[2] Kinsey tends to criticize other studies on grounds that sometimes amount to nothing more than that their methods differ from his. For example, he asserts that Terman's data on 1,242 married couples "would have been more reliable if they had been obtained by direct interviewing" (p. 31), though he neither gives nor cites evidence bearing on the issue, either in discussing Terman's and other questionnaire studies (p. 31) or in discussing his own technique of direct interviewing (Chap. 2). On p. 11 he does assert that "during the first year the value of personal interviewing as opposed to the questionnaire technique was subjected to some testing," but I have found no other reference to this testing or its results. Again, on p. 42 it is asserted, in

effect, that "things . . . can be done in a person-to-person, guided interview . . . that can never be done through a written questionnaire, or even through a directed interview in which the questions are formalized and the confines of the investigation strictly limited." Despite the lack of supporting evidence, this seems plausible to me; but so would it if its meaning were reversed by interchanging "person-to-person guided interviews" with "written questionnaire" or with "directed interview," and so does the following, from p. 62: "Whether the techniques which have been used in the present study would be equally effective with other persons engaged in studying other problems, is a question which must be answered empirically by each investigator in connection with his own special problems."

[3] The number of individuals involved in the study is given on p. 10 as 12,214. On p. 5, however, there is an outline map of the United States with the legend "Sources of histories. One dot represents 50 cases." The map contains 427 dots, so presumably represents 21,350 cases. Even if for each state one of the dots represents fewer than fifty cases, 427 dots would represent at least 19,000 cases. Kinsey uses "histories" and "cases" with different meanings, so their apparent equivalence on this map is confusing: "history" refers to the data for one individual, and "case" to the data for one individual during one five-year period of his life. The largest number that I have seen mentioned in the book as the total number of cases is 14,084 (p. 220). These are presumably derived from the 5,300 white males to whom the data in the book are said to relate (pp. vii, 6). The ratio of cases to histories for the white males thus appears to be 14,084:5,300 or 2.66. Application of this ratio to all 12,214 histories suggests over 32,000 cases, which is as much too high as the number of histories is too low to account for the 427 dots on the map.

Actually, it is not quite clear that the total number of histories is 12,214, as stated on p. 10. The histories are said to include 5,300 white males, about 1,000 non-white males (p. 6), and 5,800 females (p. 29). The discrepancy of 114 is presumably due to rounding to hundreds, although the total number of males is given at least once as "about 6,300" (p. 6) and at least once as "6,200" (p. 29). However, 12,214 is the total number shown both in the table distributing histories by year of collection (p. 10) and in the one distributing them by interviewers (p. 11).

That 5,300 is the number of "white males who have provided the data for the present publication" (p. 6) is not confirmed by any of the statistical tables in the book. The largest total I have noticed (often the totals are not shown but have to be computed) in any of the tables that appear to cover all of the white males is the 4,120 shown distributed by religion in Table 41, p. 208. This same table shows 4,940 males distributed by occupation, but since the adjacent column which is said to distribute 179 males by occupation totals 237, it may be that some individuals are classified under more than one occupation. Indeed, it may be that the 4,120 include some classified under more than one religion, for 4,102 is the number in the same table classified by education, and 4,069 is the number classified by age at onset of adolescence. Tables 37 and 41 both include what appears to be the same distribution by age at onset of adolescence, but the frequencies differ: Table 37 includes 521 more males (about one-eighth more) than Table 41, and these additional 521 have a distinctly different distribution by age of adoles-

cence than the 4,069, relatively more (37.0 per cent instead of 30.8 per cent) being below 13 years and relatively fewer (30.7 per cent instead of 36.3 per cent) above 13. Table 67, p. 298, shows 4,606 males distributed by age at onset of adolescence through 16 years, which is 43 or 44 more than shown in Table 37 for the same ages; adding to these 4,606 the 28 shown in Table 37 as reaching adolescence after 16 would give a total of 4,634 males. Table 36, p. 186 shows only 3,730 cases distributed by school grade at adolescence, and Table 35, p. 184, shows totals ranging from 1,355 to 3,573 cases for five distributions by age of occurrence of various developments in adolescence. The number for whom information is not available on a given item is never shown. In general, very little is revealed in the statistical data about the number of males covered in the volume.

The following "definition" of cases is said to "have been applied . . . throughout the present volume" (p. 682): "*Cases.* Showing the size of the population on which the data in the tables are based" (p. 683). Table 44, p. 220, shows 14,084 cases distributed by age from adolescence to 85, of whom 11,467 are 30 or under. Table 40, p. 198, shows 14,083 as the number of cases for "all ages, adolescence to 85," also with 11,467 of them 30 or under. Table 41, p. 208, however, shows the number of cases from adolescence through 30 as 11,985; so perhaps 518 cases should be added to the 14,084 in trying to discover the total number of cases involved in the study. Tables 104 and 105 (pp. 410 and 412) both show 13,359 as the total number of cases, 11,314 single and 2,045 married; 9,286 aged under 33 and 4,073 aged 33 or over. (Incidentally, this is one of the few pairs of tables I have had occasion to compare that have not proved inconsistent in their totals; the pair on pages 10 and 11 is another exception, and Tables 40 and 44 miss by only one case. Tables 104 and 105, though consistent with one another, show fewer cases 32 years of age and under than are shown as 30 and under in Tables 40, 41, and 44.) Tables 152–154, pp. 686–734, the so-called "Clinical Tables," include 15,746 cases (11,725 single, 3,275 married and 746 previously married), according to my addition. The numbers of cases shown in these clinical tables are hard to reconcile with one another, however, for the sum of the numbers shown in various subdivisions sometimes exceeds, and sometimes is exceeded by, the number shown for the whole group. For example, the data on pp. 688–690 for single white males, age group adolescence through 15 years, educational level 13+, show the whole group as 2,799 cases; but the sum of corresponding figures for the urban (2,587) and the rural (352) subdivisions of the group is 2,939, and the sum of the corresponding figures for the six religious subdivisions of the same group is 2,974. Such a result, where the whole is less than the sum of its parts, could occur if, for example, an individual who shifted from one subdivision to another were classified in both for the five-year interval in which the shift occurred—though this practice would be unsound. On the other hand, the whole frequently turns out to exceed the sum of its parts; for example, for the same age, color, and marital status just discussed but educational level 9–12, the total number of cases for the whole group is 606, but the urban (459) and rural (124) figures total only 583.

The term "population" in the definition of cases quoted above seems actually to mean "sample." The rest of the paragraph of which the quotation is

the first sentence deals with the adequacy of samples of various sizes. In general, Kinsey seems to be aware of the distinction between the statistical concepts of "sample" and "population" and of the fundamental importance of the distinction (see, for example, the section on "The Taxonomic Approach," pp. 16–21), but he misuses the terms in a way that is frequently disconcerting. For example, the heading "population in sample" appears in various tables (e.g., p. 190) and "sample population" in others (p. 188). Sometimes the phrases "sample population" and "U. S. population" (p. 188) seem to represent the sample-population distinction. The phrase "the whole population involved in the present study" (p. 194) seems to mean "the whole sample."

"Educational level 13+" is referred to above in discussing the numbers of cases shown in Tables 152–154. To Kinsey "13+" means "ultimately more than 12 years," i.e., that at least a start has been *or ultimately will be* made to college. In using the tables as a basis for comparison of any particular male, it is necessary either to exclude "persons who are still in school, since there is no certainty how far they will go before they finally terminate their education" (p. 331), or to "predict, on the basis of his home background, the amount of his future schooling" (p. 682); and presumably similar requirements apply to those who have left school before reaching the highest educational level and who might later resume their education.

It is doubtful that the age division referred to above in connection with Tables 104 and 105, and also involved in Tables 98–103, is between "persons who were 33 years of age or older at the time they contributed their histories" and "persons who were younger than 33 at the time of contributing," as stated on p. 395, since Tables 98–100 include data on the behavior of the younger group at age 33. Perhaps the younger group was actually 33 and under and the older group over 33.

[4] Five hundred and twenty-one is said to be the number of "items which are systematically covered on each of the histories in the present study" (p. 32), but this is perhaps an exaggeration, partly because the count depends on what is considered an item and what is considered merely one of several possible responses for an item, and partly because for any individual many of the items are inapplicable.

The list of "Items Covered on Sex Histories" (pp. 63–70) shows nine major groupings under which are a total of 71 numbered headings. The items listed under some headings call for distinct pieces of information; for example, under "educational history" are listed years of schooling, colleges attended, college majors, age upon leaving school, and age while in high school. The items listed under other headings, on the other hand, constitute little more than a list of possible answers to a single question; for example, under "recreational interests" are listed, among others, moving pictures, dancing, cards, hunting, fishing, reading, sewing, music, sports. There is no way to tell whether the claim of 521 includes items like those under recreation on a par with those under education.

The "actual number of items covered in each case" is described on p. 63 as "usually nearer 300, and the number involved in the histories of younger and less experienced individuals is often less than that." This suggests that 300 *or less* is the number of items involved for any one person. The following

passage from p. 50 suggests, however, that 300 *or more* is the number for each individual: "On each history in the present study there has been a systematic coverage of a basic minimum of about 300 items. This minimum is expanded for persons who have extended experience. . . . The maximum history covers 521 items." On the other hand, it appears from p. 51 that numerous items beyond the list of 521 may be covered in some histories, e.g., males with elaborate techniques of masturbation, individuals who have had some complex relation with their parents, identical twins, highly intelligent individuals with considerable experience in a socially taboo type of behavior, individuals involved in masochism or sadism, persons who are handicapped, have lived in foreign countries, have had experience in military groups, etc. "As scientific explorers, we, in the present study, have been unlimited in our search to find out what people do sexually" (p. 51).

Kinsey's comparison of his study with nineteen others in regard to number of items covered is a little misleading (pp. 28–29). For example, Ramsey's 218 items are described as 41.9 per cent as many as Kinsey's; but for boys from Junior High Schools, Y.M.C.A.'s, and Boys' Clubs, the group covered by Ramsey, Kinsey's list of relevant questions may be not much larger than 218—certainly it is not nearly two and one half times as large, as is implied by the 41.9 per cent figure. Thus what at first appears to be a difference in the amount of information for each individual is at least partly only a reflection of the difference in the number and variety of individuals covered. In fact, we learn on p. 30 that Ramsey's work was "based on personal interviews which were coordinated with the list of questions and the techniques of the present study."

[5] If we accept the implications of Kinsey's assertions (p. 199) that "there is, inevitably, some correlation between these rates [orgasms per week] and the positions which these persons take in a public debate" on sex instruction and administrative policies of educational institutions, that "the policies that ultimately come out of such meetings [on juvenile delinquency, law enforcement, and sex laws] would reflect the attitudes and sexual experience of the most vocal members of the group, rather than an intelligently thought-out program established on objectively accumulated data," and that "often the conclusions [of scientific discussions of sex] are limited by the personal experience of the author," then many of his own conclusions and implications—especially those not based on the statistical data—require for their evaluation some indication of where he himself falls in his various distributions. Acceptance of this line of argument might "explain away" many of Kinsey's social and ethical judgments, as, for example, if he were one of the six persons described on p. 217 as having the highest long-time averages.

[6] The first omission from this quotation describes in more detail the social contacts used to collect the supplementary data: visits to the subjects' homes; visits with them to their friends' homes, theatres, concerts, night clubs, taverns, and other places of recreation; correspondence; records of their sexual activities; photographs of their drawings; transcripts of their court, institutional, or social agency records. This impressive account of Kinsey's own collection of supplementary data should be contrasted with the following: "Sometimes social scientists hobnob as tourists in some social milieu sufficiently removed from their own to make it possible for them to

acquire 'impressions' and 'hunches' about 'social patterns' and 'motivations of behavior' in whole cultures. . . . The day seems overdue when scientists studying human material will forsake barbershop techniques . . . " (p. 19). This brings to mind Alfred Marshall's "general rule that in discussions on method and scope, a man is nearly sure to be right when affirming the usefulness of his own procedure, and wrong when denying that of others" (*Principles of Economics*, Eighth Edition, p. 771). Another example is provided by comparing the passage on p. 201 objecting to conclusions about sexual behavior based on persons who come to a clinic, with the passage on p. 37 asserting that while many of Kinsey's subjects were obtained because of public knowledge of the project as a source of help with personal sexual problems, these were the everyday sexual problems of the average individual.

[7] Actually, it is hard to see how even the supplementary data could support such apparently quantitative assertions as "most of the tragedies that develop out of sexual activities are products of this conflict between the attitudes of different social levels" (p. 385), or "most of the complications which are observable in sexual histories are the result of society's reactions when it obtains knowledge of an individual's behavior, or the individual's fear of how society would react if he were discovered" (p. 202).

[8] Many of Kinsey's methodological assertions are of considerable interest, so it is regrettable that he has not indicated their basis. Examples are: "Appreciation must be sincere, else it will not work" (p. 37). "The underworld requires only a gesture of honest friendship before it is ready to admit one as a friend, and to give histories 'because you are my friend' " (p. 36). "The experienced interviewer knows when he has established a sufficient rapport to obtain an honest record, in the same way that the subject knows that he can give that honest record to the interviewer" (p. 43). "People understand each other when they look directly at each other" (p. 48). "We attempted to follow standard practice [of making records only after the subject has left at the close of an interview] early in this study and found that it introduced a tremendous error into the records" (p. 50). "When one is dealing with such a socially involved question as sex it becomes particularly important to ask direct questions. . . . Euphemisms should not be used . . . " (p. 53). " . . . we always begin by asking *when* they first engaged in such activity. . . . It might be thought that this approach would bias the answer, but there is no indication that we get false admissions . . . " (p. 53). "Looking an individual squarely in the eye, and firing questions at him with maximum speed, are two of the best guarantees against exaggeration" (p. 54). Also, see Notes 2 and 15.

The correctness of these assertions is not in question here. They are cited simply as illustrations of propositions asserted without the evidence necessary to enable another investigator to evaluate them. In this sense they differ from Kinsey's conclusions about male sexual behavior, insofar as these are accompanied by statistical data whose source is explained.

[9] See Chester I. Barnard, *The Functions of the Executive* (Harvard University Press, 1938, pp. xvi+334) and *Organization and Management* (Harvard University Press, 1948, pp. xi+244); Robert Redfield, *Tepoztlan, A Mexican Village: A Study of Folk Life* (University of Chicago Press, 1930, pp. xi

+247) and *The Folk Culture of Yucatan* (University of Chicago Press, 1941, pp. xxiii+416).

In quality, as contrasted with type, Kinsey's work is not comparable with Barnard's or Redfield's. Barnard, for example, seems to have a deeper understanding of the inconsistency between legalistic statements of the principles of administrative authority in a social organization and actual behavior in the same organization, than Kinsey does of the inconsistency between our "publicly pretended code of morals" (p. 197) or our "socially pretended custom" (p. 203) and his finding that the persons involved in "illicit activities, each performance of which is punishable as a crime under the law, . . . constitute more than 95 per cent of the total male population" (p. 392).

Incidentally, this 95 per cent figure needs to be interpreted with more care than Kinsey uses. It means that 95 per cent of all white males either have engaged at least once in their lives in an "illicit" activity or can be expected to engage in one at least once, if they live to be 85. (Actually the same figure results if we assume only that they will live to be 45.) Several of the statements Kinsey makes in connection with this 95 per cent figure imply that 95 per cent of the total male population *has* engaged in illicit activities. This is logically equivalent to interpreting "all males now living will ultimately die" to mean "all males now living have already died."

No source is cited for the figure 95 per cent and it cannot be verified from the statistical data given in the book. Even if it is correct and were correctly interpreted, to conclude from it that "only a relatively small proportion of the males who are sent to penal institutions for sex offenses have been involved in behavior which is materially different from the behavior of most of the males in the population" (p. 392) would be a non-sequitur. As Terman says, "it is as though one said that if 95 per cent of all males have at some time in their lives stolen something, those who are sent to penal institutions for theft or burglary are not materially different from most males in the population." (Lewis M. Terman, "Kinsey's 'Sexual Behavior in the Human Male': Some Comments and Criticisms," *Psychological Bulletin*, vol. 45, 1948, pp. 443–459; the quotation is from p. 456.)

Kinsey makes a similar statement about adolescents: "On a specific calculation of our data, it may be stated that at least 85 per cent of the younger male population could be convicted as sex offenders if law enforcement officials were as efficient as most people expect them to be. The stray boy who is caught and brought before a court may not be different from most of his fellows, but the public, not knowing of the near universality of adolescent sexual activity, heaps the penalty for the whole group upon the shoulders of the one boy who happens to be apprehended" (p. 224).

[10] Terman, in the review cited in Note 9, makes a number of sound criticisms of the methods of obtaining data from a given subject. One that must receive attention in any consideration of Kinsey's statistics is his use of data based on long-distance memory. Each individual is asked about his activities as far back as he can recall, and his reports are tabulated by five year age intervals. Thus each individual provides several cases. "In the computation of mean frequency of masturbation at age 15, for example, the memory report of a 50-year-old counts as heavily as the report of a 15-year-old" (Terman,

p. 446). One important consequence of this, aside from inaccuracies that it may introduce, is that many of the "cases" on which the statistical analysis is based are not independent. This might explain, for example, at least in part, Kinsey's finding that patterns of sexual behavior are remarkably constant throughout life, assuming in early childhood the pattern of the occupational and educational level to which the individual ultimately will belong (p. 419). It also invalidates, at least partially, Kinsey's notion that "smooth trends in such curves are evidence of their approach to reality" (p. 132), for most of the smooth trends shown are age trends and therefore involve many of the same individuals in the successive age groups. In fact most of the charts to which Kinsey has added smoothed curves are accumulative incidence curves; not only are the successive age groups not independent here, but the calculations for any age group involve cumulating cases over other age groups.

Incidentally, the fact that some 62 of the 173 charts contain smoothed curves is surprising in view of the statement that "All of the frequency curves in this volume are based on the actual calculations, and in no instance have they been smoothed by any process or approximated by interpolations or other sorts of estimates or predictions" (p. 111). Most of the 62 smoothed curves are accumulative incidence curves, which show per cent of total population having had a specified type of experience—a sort of cumulative frequency adjusted for exposure-to-risk—plotted against age, so perhaps are not regarded as "frequency" curves. Figures 36 and 37, however, both show frequency data that have been smoothed by cumulating and further by use of a curve. Personally, however, I have no complaint against the smoothed curves, since all charts show the actual observations clearly. The simple frequency polygons, showing per cent of cases in various class intervals by average number of orgasms per week, are nearly all too erratic to permit of useful smooth curves; they thus fail to provide whatever confidence in the data would be lent by smoothness.

[11] The criteria by which cases are classified and their numbers of categories are: sex, 2; race-cultural group, 11; marital status, 3; age, 18; age at adolescence, 6; educational level, 9; occupational class, 10; occupational class of parent, 10; rural-urban background, 5; religious group, 3; religious adherence, 4; geographic origin, unspecified. The product of the 11 category numbers is 384,912,000. If Kinsey intends to represent geographic origin by 48 states, which are the only geographic units he mentions, there will be nearly 20 billion categories. So far, he has made no classifications by geographic origin, but he says "state of residence for the most continuous period of time, and the place of residence during the childhood and adolescent years, will probably represent the most significant part of the data" (p. 81). It seems to me that a few broad regions might suffice; five regions would result in "nearly two billion" segments.

[12] Quinn McNemar cites the following four fallacies in Kinsey's efforts to determine a proper size of sample: "(1) Failure to recognize the fact that the sampling stabilities of means, medians, and modes are not a function of their magnitudes, but rather of trait variability. . . . (2) Failure to consider the fact that these statistics differ markedly from each other in their sampling errors. . . . (3) Failure to observe the fact that the sampling stability of

percentages is not a linear function of their magnitudes but rather of their degree of remoteness from 50 per cent. . . . (4) Failure to note that convergence of sub-sample values to total group values must be more rapid when sub-samples are drawn from small (finite) groups than when drawn from larger groups. . . . In brief, incognizance of four elementary statistical principles renders worthless this elaborate effort to determine how large *N* should be for a sub-group" (quoted on pp. 450–451 of the Terman review cited in Note 9; the positive part of McNemar's third point seems to be an imprecise allusion to the fact that the sampling variance of a proportion is a linear function of the squared difference between the proportion and one-half.)

Such incompetence on this and other statistical points is surprising in view of Kinsey's statement, widely cited among statisticians, that "the statistical set-up of the research was originally checked by Dr. Lowell Reed of the School of Hygiene and Public Health at The Johns Hopkins University. A long list of persons experienced in sampling and in other aspects of statistics has been constantly available for consultation" (p. viii). At the Cleveland meeting where this paper was presented, Helen M. Walker, who presided, read a letter from Reed to Kinsey dated December 10, 1948 which said, in part, "I have been troubled at the flood of criticism that has been leveled at your work by the statisticians, mainly because I know that there is value in your work, but secondarily because you included my name in the preface with the implication that I had some responsibility for the analysis. As you of course remember, I saw your work only on the occasion of a two-day visit to Bloomington in December, 1942. On the basis of that visit, I joined heartily with the Committee in recommending to the National Research Council that support be continued, but a part of the recommendation was that appropriate arrangements within the budget should be made to strengthen the project from the statistical side. I became so busy with work connected with the war that I lost all contact with the project and I don't know what was done, if anything, to carry out this recommendation. If the type of statistical guidance had been provided that I had in mind, I feel sure that you would now be free of some of the criticism that is now being justly leveled at the work. . . . As you know, . . . I have seen nothing of the work between that two-day visit in 1942 and the public appearance of the book." (Transcribed from recording made at the meeting by Chester I. Bliss.) As for the "list of persons experienced in sampling and in other aspects of statistics," it would appear that either they were not consulted or else their advice was not followed.

[13] Stein's sequential estimation procedure was presented before the Institute of Mathematical Statistics at Madison on 10 September 1948, but has not otherwise been published. It is a method of determining, to prescribed accuracy with prescribed confidence, the mean of a normal distribution whose standard deviation is unknown. On the average, it requires hardly any more observations than would be required in standard single sampling if the standard deviation were known.

[14] It would be interesting to have the data for the cases in a given age group classified by the age of the subject at the time of the interview. Any systematic variation with age at interview could, however, be interpreted as reflecting either memory effects or time trends; but independence between

behavior reported for a given age and age at interview would be necessary to confirm both Kinsey's handling of cases and his conclusion that generation-to-generation changes are minor (pp. 394–417). Similarly, a study of the intra-class correlation in clusters of the kind Kinsey uses would be valuable.

[15] Kinsey's discussion of his sample is inadequate in two crucial respects: he does not explain adequately his procedure of selecting cases, and he does not describe adequately the determinable characteristics of the sample that he got.

With respect to sampling procedures, he says little except that randomness is difficult to achieve. He does not suggest that he made any efforts to approximate it; instead, he says "since it is impossible to secure a strictly randomized sample, the best substitute is to secure one hundred per cent of the persons in each social unit from which the sample is drawn" (p. 93). No basis for the assertion is given, but probably the hope is to include not only the most willing but also the more reluctant subjects. Whatever gain there might be in this respect, however, would be offset at least partially by the fact that efforts to get all of a group were not ordinarily made unless at least half had already been obtained (p. 95); thus, those groups containing relatively many reluctant subjects would be less likely to be the objects of hundred per cent drives than those containing relatively few.

As to the composition of the sample actually secured, it is hard to learn much about this; even its size is uncertain (see Note 3). Scattered through the book are various scraps of information about special groups that have been sampled. Terman, on p. 447 of the review cited in Note 9, lists some of these: Of the 62 hundred-per cent groups, 42 were of college level and 7 were delinquents or inmates of penal and mental institutions (p. 95.) "Perhaps half" of the histories were obtained through contacts resulting from lectures (p. 38). Seventeen penal or correction institutions have provided histories (p. 15). Five underworld communities and five homosexual communities are represented—they are listed as "social or civic organizations" (p. 16). There are data on 1,200 persons convicted of sex offenses (p. 392). In addition, a passage on p. 38 strongly suggests to me that "several hundred psychoanalysts, psychiatrists, physicians, clinical psychologists, social workers, and other professional persons [who] have had an especial interest in observing the interviewing techniques" were included; and this interpretation is reinforced by the section on "The Confidence of the Record" (pp. 44–47). On pp. 14–15 we learn that the sample includes persons who have been students at 528 colleges, and that 14 of these have contributed 100 or more histories apiece; even assuming that each of the 14 has contributed only 100 and that each of the other 514 has contributed only one, this accounts for 1,914 histories, or nearly 16 per cent of the total number (12,214). The map on p. 5 showing the source of the data (see Note 3) has 191, or 45 per cent, of its dots in Indiana and the four adjoining states, Illinois, Kentucky, Michigan, and Ohio—five states which contain about 20 per cent of the U. S. population (19.4 per cent in 1946).

Table 41, p. 208, which was discussed in Note 3 above, comes as close as any to revealing the characteristics of the sample. It suggests that about 60 per cent of the histories are college level (in fact, that 26 per cent have

training beyond college); that 1.6 per cent are from the underworld, 0.5 per cent are business executives, and 61 per cent are white collar or professional workers; that 76 per cent are Protestants, 12 per cent Catholics, and 12 per cent Jews; that 74 per cent are religiously inactive. What we clearly should have, however, is a definite statement of the distribution of the histories among the 163 segments for which conclusions are drawn.

[16] An example of a table that is difficult to understand is Table 14, "Comparisons of data obtained from spouses" (p. 126). The second column in this table is headed "items involved" and the third column is headed "unit of measurement." The seventh and eighth columns are headed "mean of husband's reports" and "mean of wife's reports." For the item "pre-marital acquaint." the unit of measurement is "12 mon." and the means of husband's and wife's reports are 42.11 and 40.88. For "engagement" the unit is "4 mon." and the means are 12.64 and 12.85. For "lapse, marr.—first birth" the unit is "6 mon." and the means are 28.05 and 28.19. These figures seem to say that the couples in the comparison were, on the average, acquainted 41 to 42 years before marriage, engaged over 4 years, and married 14 years before their first child was born! The data seem more plausible, however, if we assume that the units of measurement are one month in all three cases. The "units" given seem only to describe the amount of discrepancy which, if not exceeded, is called zero, i.e., identical response for both spouses. The fourth column is headed "ident. rspns. %."

Tables 152–154 are puzzling because, as mentioned in Note 3, the numbers of cases in subgroups add up sometimes to more and sometimes to less than the number of cases shown for the whole group.

[17] Kinsey says that the formula for the median is $(n+1)/2$ (p. 113). The text following this formula, however, gives a correct explanation of the median. With respect to the arithmetic mean, his formula (p. 112) is correct, except that he does not define its symbols, but it is followed by three erroneous assertions: first, that "a mean represents the total number of measurements . . . in each group divided by the number of individuals in the group"—he means, of course not the total number but the sum of the measurements, and an illustrative example included parenthetically at the omitted part of the quotation is correctly handled; second, that "the mean represents the midpoint of the measurements"—the median is the midpoint either in the sense that as many observations lie above as below it, and the mean in the sense that the sum of the deviations is as great above it as below it, but neither the mean nor the median is midway between the extremes; third, that the mean's "position . . . is therefore [i.e., because it "represents the midpoints of the measurements"] materially affected by the presence of even a few high-rating individuals in a population [i.e., sample—see Note 3]; and . . . a few high-rating individuals affect the means more than a large population [i.e. number] of low-rating individuals"—which suggest that Kinsey really believes the second of his erroneous assertions.

Kinsey also asserts that "where most of the individuals in a sample belong in a frequency class which is midway between the extremes of the distribution, and where an equal number of individuals lie in symmetrical distribution on either side of the midpoint, the mean becomes identical with the median" (p. 113). It is not clear whether this means that *either* of the two

conditions will bring about the identity, or that *both* are necessary to bring it about. Actually, the first condition is irrelevant; and the second condition is sufficient but not necessary for the identity.

[18] As a simple illustration of the use of the "U. S. Corrections," consider Table 29, "Continuity of pre-adolescent sex play with adolescent activity" (p. 174). The first section of this table shows "% with continuity" for three educational levels, as follows: level 0–8, 77.0; level 9–12, 67.4; level 13+, 29.8. Since the sample sizes ("cases") for these levels are 243, 221, and 763, the per cent for all three levels pooled into a single sample of 1,227 would be 45.9. Obviously, however, an over-all average should weight the three levels not in proportion to their sample sizes as pooling does, but in proportion to their population sizes. These population weights are given for 1940 as 53.21, 35.13, and 10.39, respectively, with 1.2 per cent of the population not reporting educational level (Table 11, p. 108). Combining the three observed percentages with these weights gives an over-all percentage of 68.6. Kinsey shows 64.9, however. Similarly, for the second and third sections of the table my calculations give 60.7 and 44.8, but Kinsey shows 54.7 and 42.1. In a similar check of the three largest percentages shown in the final column of Table 38 (p. 190), I find 68.24, 14.42, and 11.29 where Kinsey shows 68.39, 12.53, and 13.11. Since Kinsey assures us that "all mathematical calculations on this project have been performed twice, independently by each of two persons" (p. 109), I assume that I have not understood the "U. S. Corrections" correctly.

A curious feature of the "Tables for U. S. Corrections" (pp. 106–108) is that the age groups for which the weights are shown are not the same as those used throughout the report, but are a year lower, e.g., 15–19 and 20–24 instead of 16–20 and 21–25. This does not affect the discrepancies mentioned in the preceding paragraph, however.

[19] Terman's review is cited more specifically in Note 9. Another review which makes several sound criticisms of the interpretation is that by Jacob Goldstein and Nicholas Pastore, "Sexual Behavior of the American Male: A Special Review of the Kinsey Report," *Journal of Psychology*, 26 (1948), pp. 347–362. Both of these reviews have been helpful in preparing this paper, which has also benefited from critical readings by Milton Friedman and by L. J. Savage.

[20] The quotations from the *Encyclopaedia of the Social Sciences* are from J. F. Rees, "Rogers, James Edwin Thorold (1823–90)," vol. 13, p. 417, and Willard L. Thorp and George R. Taylor, "Prices," vol. 12, p. 377.

THE CITY BLOCK AS A UNIT FOR RECORDING AND ANALYZING URBAN DATA

EDWARD B. OLDS, *Research Director*
Social Planning Council, St. Louis, Missouri

Tabulations by city blocks make possible many uses of small area data beyond those which can be made from census tract tabulations. Block data can be economically analyzed and summarized by the use of summary punched cards. Some uses of block data are illustrated from the St. Louis experience. Suggestions are presented for new census data needed by blocks. The publication of a local block map and street directory facilitates the compilation of new data to supplement those obtained from the decennial census. Despite the many uses and advantages of block data, they do not replace census tract tabulations which meet a somewhat different need.

THE NEED FOR some type of standard unit geographic area for recording data about a city is apparent to many research workers and administrators faced with the problem of drawing conclusions from statistics about the city. The census tract, popularized by Dr. Walter Laidlaw and later by Howard Whipple Green, has served as the most generally accepted statistical unit area for American cities. It was developed as a compromise device to facilitate the analysis of population trends and characteristics in sections of cities. Honest attempts have been made to define census tract boundaries so that they include territory with reasonably homogeneous characteristics and with a population of from 3,000 to 6,000 persons. Unfortunately, characteristics change and what were once good boundary lines in terms of economic or ethnic indicators are not always good boundary lines. To preserve comparability from one census to another, it is necessary to keep census tract boundaries intact except for changes in city limits. However, within limits, the census tract served to reveal gross average differences between major sections of cities, as well as trends from one census to another. The census tract has admittedly an important place in the analysis of urban data, since it is small enough to show up differences between major sections of cities, and yet large enough to be easily manipulated without considerable expense. In large cities, such as New York and Chicago, it has been found necessary to combine census tracts into statistical or community areas providing more adequate bases for the computation of death rates and simplifying the mechanics of presenting and interpreting data about sections of the city.

Just as the magnifying glass is not completely replaced in usefulness by the microscope, so the usefulness of the census tract is not displaced by a more minute and precise unit area. But there are many uses of spatially ordered data about the city which can be made when they are available by a smaller unit area than the census tract. If it is conceded that a smaller unit area is desirable, the question comes up as to how much smaller the area should be and how it should be defined. Should it be something like a precinct used in organizing elections, a "beat" in police circles, or some other arbitrary grouping of city blocks? No matter what grouping is adopted, there will always be districts which cannot be made to fit the established boundaries. Does this mean that the problem is insoluble? It can be easily answered by the adoption of the city block as a unit area. Nearly all of the following types of districts are composed of city blocks.

School districts	Campaign solicitation unit areas
Water districts	Diocese and parishes
Sewer districts	Census tracts
Health districts	Neighborhood areas
Precincts and wards	Improvement districts
Police precincts and beats	Zoning districts
Meter reading districts	Fire districts
Telephone exchange districts	Library districts
Power districts	Welfare administration districts

TAX ASSESSMENT DISTRICTS

Data compiled by blocks with totals recorded in punch cards can be economically summarized by almost any of the above districts for any city. On the other hand, it is only rarely that data tabulated by census tracts or even enumeration districts can be accurately compiled according to the above types of districts. Even though an attempt is made in establishing census tracts to follow the boundaries of various administrative districts, the problem is practically insoluble without shifting the boundaries of the districts.

The proponents of census tracts sometimes argue that if administrators will not take the trouble to change the boundaries of their districts to conform to census tract boundaries, they can do without statistical information. Such an attitude fails to take cognizance of some of the very real difficulties preventing administrative districts from being brought into congruity with the boundaries of census tracts. For example, boundaries of school districts may have to be altered as there is movement of population, to maintain the proper balance of school enrollees in each school. If one district is growing in

population while its neighbor is declining, it is obviously simpler to move the boundary of the district to correct the unbalanced enrollment situation, rather than to change the capacity of the school. Moreover, for many purposes, it is necessary to have some unit smaller than a census tract to serve as an administrative area. For example, police beats, precincts, or campaign solicitation areas must be considerably smaller than the census tract with a population usually between 3,000 and 6,000. The investment of sizeable funds in capital equipment such as telephone exchanges, power lines, sewer mains or water pipes may make it impractical to change boundaries of control areas merely to make them conform with artificial statistical areas. Unless accurate summaries can be made of the expensively compiled census information, many valuable uses are lost. Of course, in some instances it is possible to make estimates and approximations by prorating census tract data or by using overlay maps. But as business and government become more scientific, there is increasing demand for *accurate* information on which to base future plans and policies. By use of the block summary punch cards, accurate summaries can be obtained economically without excessive cost beyond the cost of coding the original data in terms of blocks. In relation to the total cost of training enumerators. conducting the canvass, designating areas, coding and tabulating, the preparation of block summary punch cards is not excessive. If a five or ten per cent increase in cost makes possible a many-fold increase in the uses of urban data, such additional costs should be justified.

THE ST. LOUIS BLOCK STATISTICS PROJECT

Some indication of the possibilities in the use of block statistics may be gained by examining the St. Louis experience. In the fall of 1945, the local committee on census enumeration areas called the Metropolitan St. Louis Census Committee, obtained the cooperation of 21 business, government, welfare and educational establishments in sponsoring a local block statistics project. This involved purchasing a deck of block summary cards for St. Louis from the U. S. Census Bureau, converting the census block numbers to those used locally for over 60 years, and publishing a Block-Street Address Directory and Map. The cost of this work was largely covered by the sale of directories, maps, and sustaining memberships. The form of the directory and map, which was published by the offset reproduction of machine listings, is indicated in Figure 1. The map location code facilitates locating a specific block on the block map. It is also used to sort and list cards in geographic sequence to improve the efficiency of

mapping. The neighborhood district name symbol and water district code were included in the directory to satisfy two agencies which assisted considerably in its compilation, the City Plan Commission and the St. Louis City Water Department.

BLOCK STREET INDEX — Page 65

House Numbers From	House Numbers To	City Block Number	Foot Note	U. S. Census Tract No	U. S. Census Block No	Neighborhood District	Map Location	Water District
MICHIGAN AVENUE EVEN								
3900	3998	2572		24A	9	MARGET	PJ37	14
4100	4198	2629		15C	18	MARGET	PJ36	14
4200	4230	2655N		15F	31	DAKOTA	PJ35	14
4232	4298	2655S		15F	30	DAKOTA	PJ35	14
4300	4398	2679		15F	29	DAKOTA	PJ35	14
4400	4498	2702		15F	28	DAKOTA	PJ34	14
4532	4598	2721		15F	27	DAKOTA	PJ34	14
[illegible]	4698	2746		15F	26	DAKOTA	PJ33	15
	4798	2759		15F	25	DAKOTA	PJ33	15
	5098	2798		15F	24	BELLER	PJ32	15
	5198	2797		15F	23	BELLER	PJ32	15

House Numbers From	House Numbers To	City Block Number	Foot Note	U. S. Census Tract No	U. S. Census Block No	Neighborhood District
MICHIGAN AVENUE ODD						
3901	3999	2573		24A	8	MAR
4101	4199	2630		15C	16	
4201	4299	2654		15F		
4301	4399	2680		15F		
4401	4459	2700		15F		
4433	4599	27[illegible]		15		
4601	4699	2745		15		
4701	4799	2760				
4901	5099	2799				
5101	5199	2800E				
5201	5299	2824				

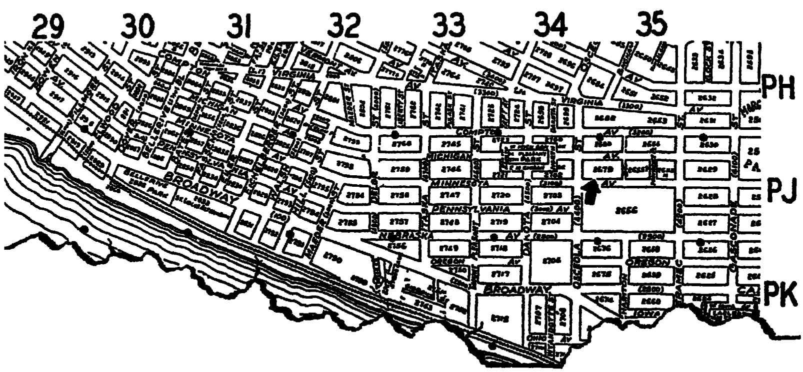


FIGURE 1. ILLUSTRATION OF FORM OF BLOCK—STREET DIRECTORY AND MAP

The use of locally established block numbers facilitates obtaining and compiling current local information. The chief data compiled regularly are:

1. Number of dwelling units in new homes for which building permits have been issued.
2. Number of dwelling units in homes for which demolition permits have been issued.
3. Number of white and Negro children enrolled in public elementary schools.

These data are useful locally to provide some indication of marked increases or decreases in population in particular neighborhoods. The school enrollment data are particularly useful in revealing annual shifts in the location of the Negro population of St. Louis. Since the

city and school authorities routinely code their records by city block numbers, the cost of making block tabulations using punch card machines is comparatively small. Summaries are tabulated by census tracts, neighborhood districts, census districts, precincts, and wards. The preparation of these summaries is facilitated by the use of a master deck of cards containing a series of code punchings signifying to which census tract, neighborhood district, etc., the particular block belongs.

For recording summary statistics about blocks, use is made of a specially printed card illustrated in Figure 2. The fields lettered from A to Q are used for quantities, such as the number of dwelling units, white school children, or dwelling units constructed in 1944. The fields printed with city block number, U. S. census tract number, block number, etc. are used for standard area codes. Reproduced decks of cards in the master file can be prepared for use by members having their own machine facilities.

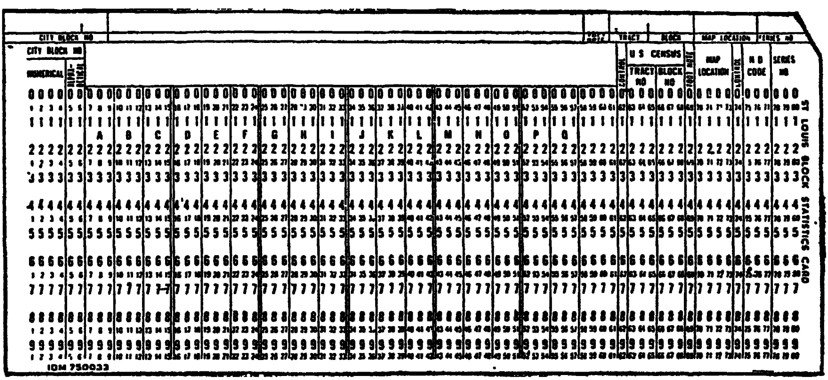


FIGURE 2. PUNCH CARD FORM USED FOR RECORDING VARIOUS TYPES OF STATISTICAL DATA FOR CITY BLOCKS

ECONOMIC RATING OF BLOCKS

To provide a convenient means of classifying addresses by economic status, a block economic rating on the basis of 1940 rents was prepared. The block summary cards from the 1940 census contained information on the average rent in each block. These cards were sorted by this average rent, listed, and at the same time the number of dwelling units in each block was cumulated. Those blocks with the lowest rents which included one per cent of the dwelling units in the city, were given a code of "01." The blocks with slightly higher rents which included another one per cent of the homes, were given a code of "02." This process

was continued until the blocks with the highest rents, which included one per cent of the homes, were given a code of "100." Addresses coded in terms of this block economic index, can be conveniently grouped into economic tenths, fifths, thirds, etc. Figure 3 illustrates the relationship between the block economic code and average rents.

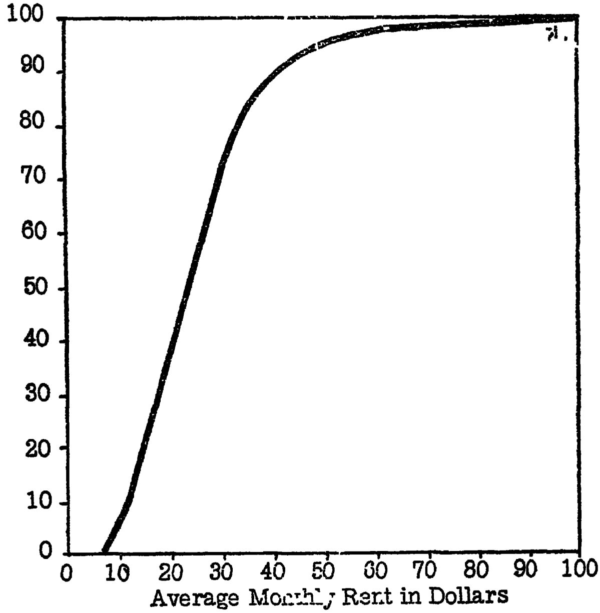


FIGURE 3. CUMULATIVE PER CENT OF HOMES IN ST. LOUIS BLOCKS WITH LESS THAN SPECIFIED AVERAGE RENTS 1940

Some of the uses which have been made of the block economic code may be of interest. In a public opinion survey, a random area sample was selected, using blocks as primary sampling units. Since it was not practical to make follow-up calls on every family in the sample, the possible biasing effect of differences in the percentage responding from low and high income areas was controlled by means of the block economic code. Tabulations were made of the number of families in the sample from five groups of economic areas classified by means of the block economic code. The distribution of usable questionnaires from these five groups of economic areas was also determined. Any significant differences in the distribution of questionnaires and the distribution of the families in the sample were corrected by obtaining more interviews. In this way, the economic composition of the families

represented by the questionnaires analyzed was kept close to the composition of the population.

In a study of subscriptions to the Community Chest obtained through neighborhood solicitation, the block economic code was used to provide an index of economic status for each solicitation area. A comparison between this economic index and subscriptions per family, indicated a marked association as illustrated in Figure 4. This information was helpful in determining the areas where neighborhood solicitation produced insufficient returns to justify the costs involved.

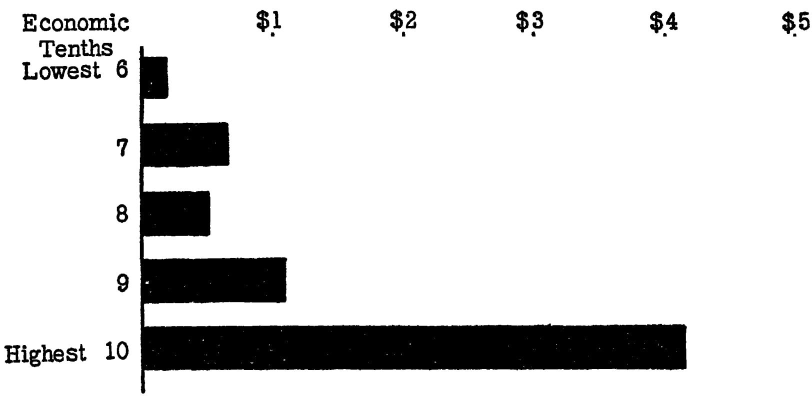


FIGURE 4. AVERAGE 1947 COMMUNITY CHEST GIFT PER FAMILY IN NEIGHBORHOOD SOLICITATION OF PART OF ST. LOUIS CLASSIFIED BY ECONOMIC TENTHS

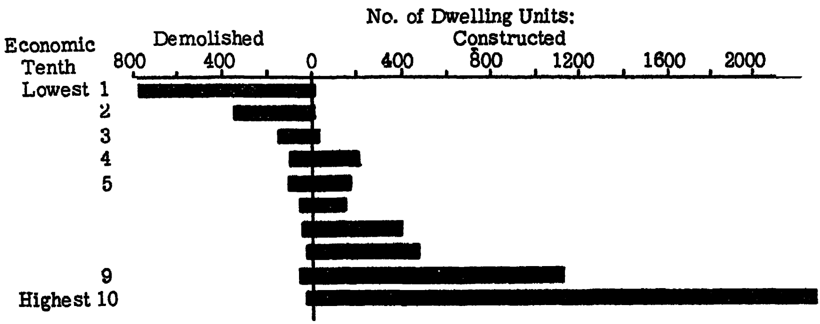


FIGURE 5. DWELLING UNITS CONSTRUCTED AND/OR DEMOLISHED UNDER PRIVATE AUSPICES IN ST. LOUIS ECONOMIC TENTHS DURING THE PERIOD 1940–46

In a study of dwelling units constructed during the period 1940 to 1946, the economic status of the blocks in which the construction took place was determined by the block economic code. A tabulation of the

units constructed and demolished according to economic tenths (each tenth containing one-tenth of the 1940 dwelling units) indicated a highly skewed distribution as shown in Figure 5. The areas with the

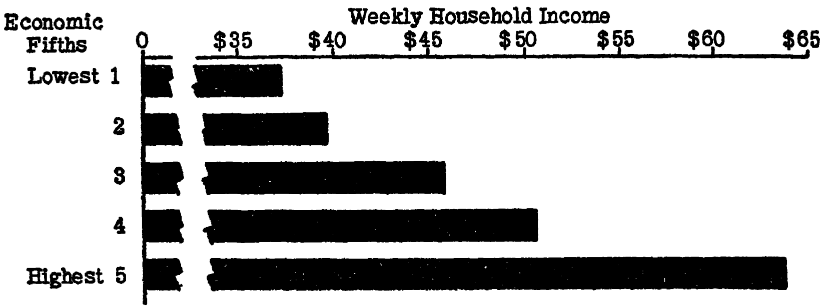


FIGURE 6. MEDIAN WEEKLY HOUSEHOLD INCOME REPORTED IN JULY 1947 BY FAMILIES IN EACH ECONOMIC FIFTH OF ST. LOUIS

highest economic status had the most building, and the areas with the lowest economic status had the least building going on. On the other hand, the high economic areas had the least demolition of homes while the low economic areas had the most demolition taking place.

Several applications of the block economic code have been made in classifying addresses in St. Louis City according to economic status. Although considerable error can result through such a method for estimating the income status of an individual, it is believed that for groups of persons or families, this method provides a fairly reliable index. Figure 6 illustrates the relationship that was found between average income reported by families in a sample survey and the block economic code based on 1940 rents. The interviews with the families were conducted in July 1947, and each family was asked to indicate in which one of the following classes their household income would fall:

Under $25 per week
$25 to $49 per week
$50 to $99 per week
$100 or more per week

A study of the economic status of the blocks in which more than 40 per cent of the population was Negro, indicated a heavy concentration of Negro blocks in the lower economic brackets. As shown in Figure 7, none of the Negro blocks were in the highest economic tenth, while 20.3 per cent of the Negro homes were in the lowest economic tenth.

Although no analysis of vital statistics data has been made, using the economic code, it is believed that highly significant differences

would be found from a comparison of life expectancy, infant deaths, etc., in low and high income areas. Such analyses require the coding of

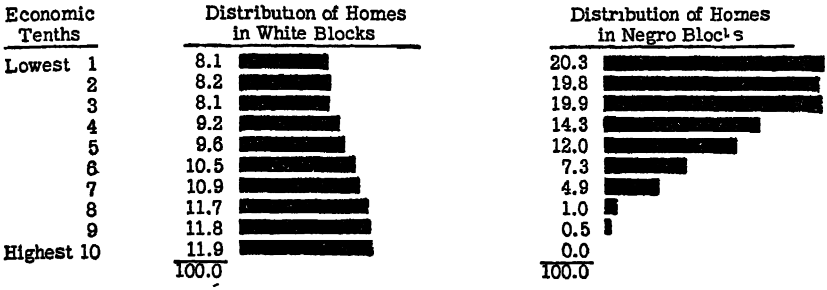


FIGURE 7. COMPARISON OF DISTRIBUTION OF HOMES IN WHITE AND NEGRO AREAS OF ST. LOUIS ACCORDING TO ECONOMIC TENTHS
Note. Each economic tenth included 10 per cent of the homes in St. Louis in 1940.

births and deaths by blocks, as well as the tabulation of population census data by blocks.

COMPILATIONS FOR SPECIAL AREAS

The block data on punch cards have been used to obtain summary tabulations of housing, school enrollment, and building permit statistics for such areas as neighborhood districts, precincts, and wards. The City Plan Commission has divided the city into 99 neighborhood and industrial districts, basing the determination of boundaries upon such factors as major streets, railroads, proximity to parks and playgrounds, land use, etc. Until block statistics were available, it was not possible to obtain accurate housing statistics for these basic planning units. As part of the St. Louis block statistics project, summary tabulations of the 1940 housing census block statistics were obtained. These included the following data:

Residential structures	Negro families
Dwelling units (homes)	Homes with more than 1.5 persons per room
Owner occupied homes	Homes needing major repairs
Tenant occupied homes	Homes without private bath
Homes built 1930 to 1939	Average monthly rent of homes
Homes built 1920 to 1929	Total rent
Homes built 1900 to 1919	Number reporting rent
Homes built before 1900	

VOTING BEHAVIOR STUDY

Summaries of housing data have been made for the precincts and wards of St. Louis. The summarized data were then used to compute

octile ratings for each precinct in each of four housing factors. This was done by computing percentages for each precinct, ranking the percentages, and then grouping the ranked precincts into eight groups. The four factors were as follows:

1. Per cent of homes owner occupied.
2. Per cent of homes built before 1900.
3. Per cent of public school enrollees who were Negro in Nov. 1946.
4. Average rents.

The block data were also used to prepare estimates of the population 21 and over in each precinct as of January 1, 1948. These estimates were based upon the current estimated number of families, using the 1940 count of dwelling units, plus units represented in building permits issued since 1940, and less dwelling units represented in permits for demolitions since 1940. The estimated number of families was multiplied by the 1940 ratio of population 21 and over, to families in the nearest census tract. The sum of these products for the city was compared with the estimated population 21 and over in the city. The provisional estimate for each precinct was then multiplied by a correction factor so that the figures finally used add up to the estimated population 21 and over in the city. While this method is subject to considerable error, it was considered more reliable than any other available method for obtaining a current estimate of population 21 and over. Percentages and octile ratings were then computed for the proportion of the voting population registered to vote. A series of 28 other octile ratings was computed from the election statistics on civic issues, as well as for political parties in eight elections held since Nov. 1944. Comparisons were not made prior to this time because of non-comparable precinct boundaries. The data prior to 1948 for each precinct were summarized on one specially printed tabulating card. Other punched cards were used to list the statistical data onto the printed card. Complete sets of 784 precinct data cards were turned over to the sponsors of the project. Other sets can be prepared economically from the master cards. The top line of each card contains a series of 23 octile ratings. The percentages upon which these ratings were based are specified by small numbers printed below each rating and in the lower left corner of each percentage cell. A set of punched cards containing the octile ratings was used for an intercorrelation analysis, using the tetrachoric correlation method. This analysis indicated the following significant relationships between the housing and voting indexes:

1. Democratic precincts tended to remain Democratic and Republican precincts tended to remain Republican.
2. Precincts with a large proportion of the population registered tended to have a large proportion of the registrants voting in each election.
3. Areas with high home ownership had a larger proportion of the population registered than areas with low home ownership.
4. High rent areas were more inclined to vote Republican than low rent areas.
5. Areas with many old homes opposed daylight saving time and a new state constitution.

TABULATION OF SCHOOL ENROLLEES

Each year in November, the Board of Education asks each elementary school to prepare a report listing the block numbers in which their pupils reside and the number of pupils in each block. The Block-Street Address Directory is used in coding addresses by blocks. Since St. Louis has a completely segregated school system, it is possible to make tabulations of these data to show the number of white and Negro school enrollees in each block. Such tabulations have been made for each of the following years:—1941, 1945, 1946, 9147. From these data by block, it has been possible to prepare a map which shows the trend of movement of the Negro areas in St. Louis during the period 1941 to 1947. Figure 8 illustrates the *Post Dispatch* map[1] drawn from the more precise block map published in two colors by the Social Planning Council and the Urban League. The housing statistics and land use data by blocks were used to compute differences in dwelling units per residential area between white and Negro areas.

ESTABLISHING CAMPAIGN DISTRICTS

Uses of the Block-Street Address Directory can be shown by describing a project involving the grouping of addresses of campaign prospects into convenient solicitation control areas. Cards were punched alphabetically giving the names and addresses of about 20,000 prospects. The punching of names and addresses was needed for the preparation of prospect lists, pledge cards, and mailing strips. In punching the addresses, house numbers were punched in one field while street names were punched in another field. Accordingly, it was possible to mechanically sort the cards by street and house number, keeping the

[1] The *St Louis Post Dispatch* carried a feature article by Richard G Baumhoff on this study in the Sunday issue, Aug 15, 1948

odd house numbers separate from the even house numbers. The use of the Block-Street Address Directory, together with a listing of these sorted address cards, privided a highly efficient means of establishing ing the block codes. In marking the list with the block code, it was found that usually from five to 20 adjacent listings would be in the same block. The punching of the block code into the cards was accomplished by manually filing the cards behind pre-punched block

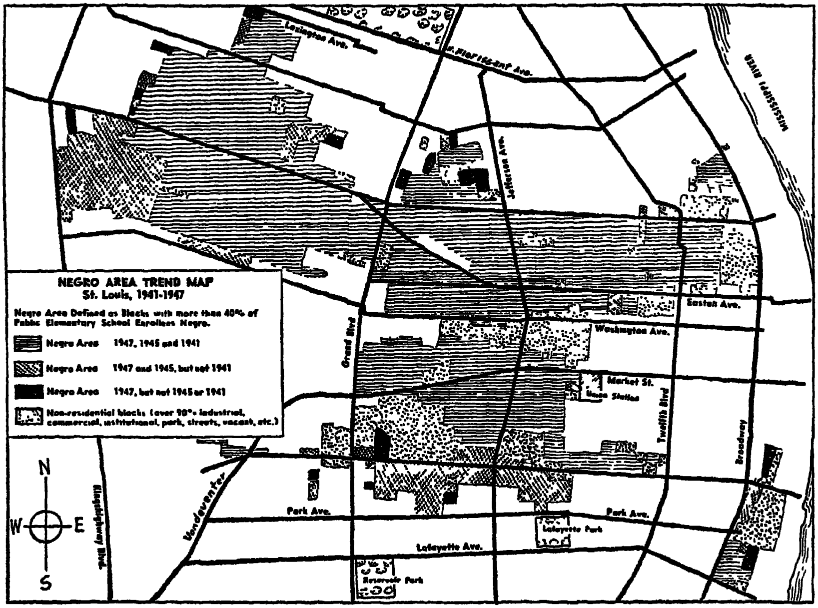


FIGURE 8. NEGRO AREA TREND MAP

The author wishes to acknowledge with thanks the permission granted by the *St. Louis Post Dispatch* to publish this map.

master cards and then intersperse gang punching the cards. The cards were then sorted down and tabulated by block. Work maps were posted with the number of prospects in each block and area boundaries were drawn to include the proper number of prospects in each area.

MAKING SPOT MAPS

In St. Louis, the city block numbers consist of four numerical digits and two alphabetical suffixes. Although this makes a rather cumbersome number, it is useful because many city records and maps are referenced with these official city block numbers. However, to locate

any given block efficiently, it is necessary to have what is called a supplementary "map location" number. This number consists of two letters followed by two numbers, like "PQ42" which defines a particular square of land in the city with sides one-fourth mile long. Every block is assigned to one specific square on the basis of where the majority of its area is located. Cards punched with city block number or census tract and block number can be automatically gang punched with this map location number and other area codes at one operation. They can then be sorted and listed in geographic columns and rows which greatly facilitates the spotting of block maps to show the accurate distribution of addresses.

POTENTIAL USES OF BLOCK DATA

The foregoing examples represent only a few of the possible uses of population and housing statistics by blocks. A consideration of these uses should suggest many others which would be made if data were available uniformly for every metropolitan area including the suburbs as well as the central city. The fact that accurate summaries can be made economically for a wide variety of administrative and study areas opens up uses which cannot be made of census tract statistics.

Some of the uses which could be made through a more general availability of block statistics are indicated in the following list:

Determination of fire, theft, and life insurance risks in different types of neighborhoods.
Studies of land values as related to population, sales, etc.
Determination of business areas of the Metropolitan District.
Planning changes in the location of transportation and utility lines.
Appraisal of property for loans or taxation.
Determination of cost of governmental and philanthropic services in each section of the city as compared to tax income obtained.
Planning optimum location of public, private, or commercial facilities for recreation, education, sales, health or welfare service, etc.
Estimates of sales or consumption using block statistics in the design of the sampling plan.
Indexing detail real estate or land use maps.
Determination of optimum districts for neighborhood improvement, police beats, meter readers, relief investigators.

SUGGESTED NEW CENSUS BLOCK DATA

One of the limitations upon the use of block statistics is the paucity of information generally available by blocks. It is believed that there could be a considerable increase in the variety of data tabulated from the decennial census without greatly increasing costs. If territory is assigned enumerators by blocks (as was done in the 1940 housing

census) the punching of a block designation in all punch cards made from the schedules would be comparatively simple. Then special tabulations could be run by blocks or groupings of blocks for any desired detail. Routinely, it is believed that certain summary information could be tabulated by blocks recording the totals in summary cards. The following items could be recorded in three decks of such cards:

Housing data (in order of importance):

Number of dwelling units.
Contract or estimated monthly rent.
Owner occupied dwelling units.
Dwelling units occupied by non-white persons.
Dwelling units built before 1900.
Dwelling units with no private bath.
Number of dwelling structures.
Dwelling units according to type of structure (3 groups).
Dwelling units without private flush toilet.
Dwelling units without running water.
Dwelling units without mechanical refrigeration.
Dwelling units without central heating.

Population data (in order of importance):

Number of persons.
Age distribution (5 groups)
Sex and color (4 groups).
Population 25 and over by years of school completed (6 groups).
Population 14 and over by employment status and sex (12 groups).
Employed persons 14 and over by major occupation group (9 groups).

These data could be economically published by listing them on plastic offset plates and reproducing several hundred copies for sale to users at a charge set to write off the publication cost. Users would be encouraged to purchase decks of cards on printed card forms clearly indicating the information punched in the cards. Users could also be supplied at cost with block maps of adequate scale for work purposes, reproduced through a blue print process from masters kept in the Census Bureau. Such maps should be in sections that would be small enough to be handled easily on normal size desks and drafting tables, and yet so made that they could be easily assembled to make up a one-piece map for a Metropolitan District. A certain amount of skilled consultation service should be made available by the Census Bureau to help users in making the best possible use of the block data.

SUGGESTED LOCAL BLOCK TABULATIONS

The preparation of block statistics as outlined above would require

local committees in each community to help in making the best uses, as well as in promoting the compilation of local material. One of the first projects for each city committee would be the purchase of decks of the summary cards and sets of block maps in the form of negative blue print masters. Another project would involve the compilation of a local block-street address directory to facilitate the compilation of local data by blocks. The promotion of local tabulations should include consideration for obtaining the following types of information:

Building erections and demolitions.
School census or school enrollment data.
Land use statistics.
Police arrests.
Juvenile delinquency cases.
Births and deaths.
Persons receiving welfare services (chronically ill, tubercular, mentally ill, general hospitalization, foster home placement, etc.).
Old age assistance, aid to dependent children, and general relief cases.
Tax assessments and collections.
Fire losses.

The financing of local projects can be handled in various ways, depending upon the community. Generally, it is possible for each administrative agency to include in its budget, small amounts sufficient to cover the processing of statistics, falling within its jurisdiction. Sales of directories can be used to cover the cost of their compilation and publication. Contributions from utilities, banks, chambers of commerce, universities, foundations, real estate firms, etc., can be used to write off the cost of local projects. However, it is necessary to have interested and competent leadership for local committees. Such persons may be found in a local city plan commission, council of social agencies, university, chamber of commerce, utility, board of education, etc. If interest warrants, it may be possible in certain communities to establish agencies equipped with staff and machinery for the most efficient processing of statistical data having general community-wide significance. Block statistics would be one of the kingpins in such a community research agency.

National agencies and concerns should find considerable uses for block statistics when they become uniformly available together with adequate maps and street address directories. Survey and polling organizations should be able to effect economies and improvements in their work through the use of block statistics. With adequate materials and interpretation, it should be possible to cover some of the added

census cost of block statistics through the sale of cards, maps, and listings.

COMPARATIVE ADVANTAGES OF CENSUS TRACT AND BLOCK DATA

From the St. Louis experience, we find that block statistics are essential for the following types of analyses:

1. Compilation of census statistics for administrative areas which are not multiples of census tracts or enumeration districts.
2. Compilation of census statistics for areas within a specified distance or travel time of a particular geographic location (half mile or mile).
3. Classification of addresses according to economic status based upon economic index computed for blocks.
4. Appraisal of neighborhood characteristics in immediate vicinity of a specified address.
5. Determination of exact boundaries of areas inhabited predominantly by a particular ethnic group.
6. Design and selection of area samples for use in factual, attitude, and opinion surveys.

Census tract data are inadequate, although better than no data, for analyses such as the above. However, census tract data are preferable to block data for analyses such as the following:

1. Community studies of districts or sections within an urban area when the population of the district is over 20,000.
2. Computation of ratios such as tuberculosis death rates for sections of a city.
3. Presentation of a general view of the variation from community to community within a city in significant population or housing characteristics.

In large cities even census tracts are too small for use in analyses such as the above.

In conclusion, statistical tabulations of selected types of data by city blocks supplement, rather than displace tabulations by census tracts. Wherever possible, administrative districts within a metropolitan area should be established as multiples of census tracts. Block statistics should be used only for analyses which require greater geographic detail than can be provided by census tract data. The judicious use of block and census tract statistics can make a noteworthy contribution to the more scientific administration of business and governmental services.

THE RELATION OF THE NET REPRODUCTION RATE TO OTHER FERTILITY MEASURES

T. J. Woofter

Recent population literature has been critical of the net reproduction rate on the grounds that it is based only on the female population, that it assumes the invariable continuation of the reproductive situation of a single year, and that it ignores the past childbearing experience of the generation.

Alternative measures are:

Male reproduction rates;

Marital reproduction rates, which are of two types—(a) those showing the birth rates of a single year standardized for duration of marriage, and (b) those showing the number of children ever born to women who have been married for varying periods of time.

Generation rates, which are based on the total number of children ever born to a generation of women who have completed the childbearing period.

Standardized quota reproduction rates are proposed in order to preserve the generation principle, but center the experience measured closer to the current year.

Rates adjusted for the order of birth of children, as proposed by Whelpton, may be calculated either from the base of a single calendar year or from the complete experience of a generation.

All of these rates may be classified into two types: (1) Those which depend on the birth rates of a single year, and (2) those which cumulate the experience of a group over a period of years. The former are more sensitive to short-time changes in the birth rate, while the latter provide a longer and more stable base for measuring trends.

I. CRITICISMS OF THE NET REPRODUCTION RATE[1]

SINCE THE FIRST exposition of the net reproduction rate and Lotka's demonstration of its relationship to the true rate of natural increase, it has been *the* measure of fertility used by demographers and has provided the techniques for some of the most penetrating analyses of fertility which have been produced in the two past decades. Recently, however, the dissatisfaction with the net reproduction rate as

[1] The most recent discussion of the net reproduction rate, which has been presented by Lotka, was contained in a paper soon to be published in the Proceedings of the Meeting of the International Statistical Institute, held in Washington in 1947. Cf. also *Journal of the American Statistical Association*, June 1936, "The Geographic Distribution of Intrinsic Natural Increase in the United States and the Examination of the Relation of Measures of Net Reproduction."

the measure of fertility, especially of fertility trends, has led to a reappraisal of the possibilities of alternative measures.

The gross reproduction rate is the sum of the age-specific female fertility rates of all women. The net rate is derived by multiplying the age-specific female fertility rates by the corresponding female survival rates and summing the products; thereby yielding a ratio of the number of daughters who will survive to the age at which their mothers gave them birth to the number of women in that age, the underlying assumption being that fertility rates or the mortality rates of a particular year will remain constant.

Some of the criticisms of the net reproduction rate seem to have been caused by a confusion as to the uses to which the rate may be put. It has, in other words, sometimes been carelessly used by those who are not familiar with all of its assumptions. This has been especially true when attempts have been made to use the customary reproduction rates for prediction. Three general uses to which such rates have been put may be distinguished:

First, they are used for the simple comparisons of two or more populations by means of rates relating to the same year, or of short-term variations within the same population. Examples are the comparison of urban and rural fertility in 1940, or the comparison of fertility in a particular area in two successive years. Obviously, the accomplishment of this purpose often requires a measure of fertility performance at a particular time which is sensitive to all of the combined factors affecting fertility. For such general purposes, the crude birth rate is often satisfactory, or the age-standardized female birth rate, which is the gross reproduction rate, may be used.

The second objective of such rates is the analysis of the factors affecting fertility, by attempting to isolate the effect of the principal factors being studied. It is especially important, when correlations of economic or social variables are made with the birth rate, that as many other factors be standardized as possible. The requirement for these purposes is a series of measures standardized for different factors in order that the difference between standardized and unstandardized rates may reveal the fluctuations which arise from the factor under consideration.

The third purpose is the endeavor to predict the future by the net reproduction rate of a particular year or of a series of years. This interpretation of the rate sometimes creeps into the thinking of the layman and even of the uninitiated investigator. This confusion is understandable, since the very term "reproduction" implies a look

into the future. The question is often asked in this way: "Is this population reproducing itself?" This is only another way of asking the question, "Will the next generation be as large as this generation?" In interpreting such a dynamic situation, some demographers have used a measure which assumes static conditions.

All through the recent European literature which criticizes the net reproduction rate we find such terms as "invariable fertility," or "relation of the tendency value to the situation of the moment," or "the long-term prospects of population growth"; the implication being that, inasmuch as the net reproduction rate does not satisfactorily provide such prediction factors, there should be a rate which does. This concept, obviously, gets into the realm of philosophical reasoning. Are there in various populations inherent fertility trends which change slowly, but persist for a long period of time? In recent discussions of fertility measures, one gets the impression that some demographers think that there are such invariable fertility trends. If we assume that there is some underlying fertility pattern, there is difficulty not only in choosing the proper measure to characterize it, but also, if prediction is attempted, the investigator is confronted with all the technical difficulties which beset extrapolation in any field.

The three principal objections which have been raised against the net reproduction rate are: (a) It applies only to females and takes no account of the sex ratio and difference in ages of fathers and mothers; (b) that, for this rate to be meaningful in practical terms, it assumes the invariable extension of the reproductive situation of a particular moment; and (c) that it ignores the past in that it does not make allowance for the influence of the past fertility experiences of the women who have children in a particular year.

Even though the recent reversal of the drop in the birth rate may be temporary, it has been sharp enough to cause a new crop of analyses of fertility and to intensify experimentation with various measures which might explain the phenomena. In the case of this country, the native white gross reproduction rate fell from about 160 in 1915 to 104 in 1936 and rebounded to 159 in 1947. This violent fluctuation is in itself concrete evidence of the small probability that the rates of any one year will remain invariable, which is a basic aassumption in the use of the net reproduction rate as a measure of generation reproduction.

Critics of the net reproduction rate argue that, while long-time trends may continue with some stability (if such long-time trends can be discerned), still the variations in economic and social conditions

and in familial attitudes change rapidly and have different effects on short-term fluctuations in fertility. Hence, they have endeavored to analyze the effect of these short-term variables and, secondarily, have groped for a measure of what might be characterized as the underlying fertility trend. The lines of approach which are treated in the following pages are: (1) Male reproduction rates; (2) nuptial reproduction rates; (3) generation reproduction rates; (4) cohort replacement rates; and (5) rates adjusted for order of birth of children.

It is not possible in the scope of one article to develop the details of each of these types of measures, but their chief characteristics may be described briefly, especially their relationship to the net reproduction rate and the availability of data for their calculation.

II. MALE NET REPRODUCTION RATES[2]

In theory, as well as in practice, the male net reproduction rate is similar to the female, except that it measures the number of sons born to 100 fathers who will survive to the age which their father had attained when they were born. It would appear that, if the population is assumed to be tending toward an equilibrium, when 100 women produce 100 surviving daughters, then it would also be tending toward an equilibrium when 100 fathers produce 100 surviving sons. If, therefore, for a particular year (as in England in 1938) the paternal net reproduction rate was .881 and the maternal only .808, which is the most pertinent as the basis for judging the effect of 1938 fertility conditions on future trends?

Difference between the mean length of male and female generations is not the only cause of discrepancies. There is a difference in the proportion married at various ages; there is a difference between the distribution within the childbearing ages of men and women; and there are differential mortality rates. All of these factors are reflected in differences between male and female fertility rates. While there is no apparent reason for choosing the rate of one sex as more useful than the rate of another, comparisons between the two are revealing as the effect of the sex ratio and differential age of marriages in the population.

---

[2] Myers, R. J., "The Validity and Significance of Male Net Reproduction Rates," *Journal of the American Statistical Association*, Vol. 36, No. 214, June 1941.

Hajnal, J., "Aspects of Recent Marriage Trends in England and Wales," *Population Studies*, Vol. 1, No. 1, June 1947.

Tietze, Christopher, "Differential Reproduction in the United States," *American Journal of Sociology*, Vol. 49, No. 3, 1943.

———, "Differential Reproduction," *Milbank Memorial Quarterly*, Vol. 19, No. 3, July 1939.

Karmel, P. H., "The Relations Between the Male and the Female Reproduction Rates," *Population Studies*, Vol. 1, No. 3, Dec. 1947.

III. NUPTIAL REPRODUCTION RATES[3]

Nuptial reproduction rates have been in use for some time, but, in recent years, experimentation with these rates by European scholars has increased to a great extent. In fact, some European demographers assert that such rates are the most satisfactory which have been evolved to date. As a result, their techniques have been considerably refined, the chief modification being a refinement to allow for duration of marriage. The previously used nuptial rates (unadjusted for duration) were constructed on the same theoretical framework as a conventional net reproduction rate, in that they were derived from age and marital specific rates based upon the fertility and mortality experience of a single year. However, they had the advantage of being specific for nuptiality and allowing the student to isolate this factor for special study. The use of such a rate as a basis for prediction is analogous to the use of the net reproduction rate, in that it is assumed to indicate the eventual rate of increase of a population with stable age structure and invariable extension of the fertility, marriage, and mortality rates of the year in question. For practical purposes, this is evidently open to the same objection which the advocates of nuptial rates have made against the conventional net reproduction rate.

Refinement of the general nuptial rate to allow for duration of marriage is made in two ways:

(1) Fertility rates according to nuptiality are calculated for each age and duration of marriage, and combined into a rate for all married women by the use of a nuptiality table. These rates are again based on a single calendar year's experience and, therefore, open to the same objection which we continually emphasize; namely, that, as a predictor, this assumes the invariable extension into the future of a single year's fertility, mortality, and marital rates, with the added objection that another factor (fertility rates by duration of marriage) is also assumed to remain constant. (For illustration, cf. Table 1.)

An ingenious refinements of such rate has recently been proposed by Hyrenius.[3] This method takes into account both the male and female age and marital distribution by duration of marriage. Its derivation is described by him as follows: "(a) The elaboration of an index of the proportion of sexes among the non-married persons within

---

[3] Glass, D. V., *Population Policies and Movements in Europe*, Appendix, pages 399–405.

Hajnal, J., "Analysis of Recent International Recovery in the Birth Rate," *Population Studies*, Vol. 1, No. 2, September 1947.

Quesnel, Carl-Erik, "Population Movements in Sweden in Recent Years," *Population Studies*, Vol. 1, No. 1, June 1947.

Hyrenius, Hannes, "La Measure de la Reproduction et de Accroisment Naturel," *Population*, April–June 1948.

certain age limits; (b) the calculation on the basis of this index and of observed nuptiality rates of a table adjusted for feminine nuptiality and the distribution of the newly married by age; (c) construction, on the basis of mortality and divorce statistics, of a table of attrition of marriages for various groups of ages at marriage; adjustment of an

TABLE 1*

FEMALE MARITAL FERTILITY RATES BY DURATION OF MARRIAGE
(Age of wife: 15–49 years)

Duration of Marriage (single years)	Fertility Rate Per 1,000 Married Women			
	1933	1939	1941	1943
0–	363	320	325	376
1–	197	198	212	239
2–	167	171	168	197
3–	144	154	151	195
4–	129	133	133	165
5–	112	123	115	150
6–	100	109	101	135
7–	89	95	90	123
8–	82	83	80	104
9–	73	74	70	90
10–	67	66	62	82
11–	59	58	56	72
12–	55	54	51	63
13–	52	48	45	54
14–	48	45	39	46
15–	40	34	35	39
16–	38	31	32	34
17–	32	30	27	30
18–	25	25	22	24
19–	26	21	19	21

* From "Population Movements in Sweden in Recent Years," by Carl-Erik Quensel, *Population Studies*, June 1947, p. 34.

analytical function to this table; (d) calculation, by combining the distribution by age of the newly married with the table of attrition of marriages, of the distribution of married women by age of the husband and duration of marriage; (e) calculation, on the basis of the preceding functions of the rate of reproduction and intrinsic rate of natural increase, legitimate fertility being given by age of the mother associated with duration of marriage, since illegitimate fertility is only known by age groups." This is, manifestly, a complex computation—one which requires more data than are available for most populations.

(2) When births are recorded, year by year, according to age of mother and duration of marriage, it is possible to trace a cohort of

marriages back to the date of marriage and construct a table showing the number of children ever born to women married in certain years, according to age and duration up to the most recent year available. A similar table can be constructed from census enumerations, which record data on duration of marriage, age, and number of children ever born (cf. reference above to Hajnal). (For illustration, cf. Table 2.)

TABLE 2*

AVERAGE NUMBER OF CHILDREN TO MARRIAGES, BY DURATION OF MARRIAGE, AT THE END OF THE YEARS 1933, 1939 AND 1943

(Age of wife at marriage, 20–24 years)

Duration of Marriage (years)	Average Number of Children Per Marriage at the End of		
	1933	1939	1943
1	0.45	0.39	0.43
2	0.68	0.62	0.66
3	0.88	0.84	0.82
4	1.06	1.01	1.06
5	1.24	1.14	1.15
6	1.42	1.27	1.33
7	1.58	1.41	1.43
8	1.73	1.53	1.60
9	1.89	1.64	1.62
10	2.03	1.74	1.71
11	2.17	1.85	1.80
12	2.36	1.97	1.88
13	2.53	2.09	1.95
14	2.71	2.20	2.03
15	2.82	2.31	2.06
16	2.96	2.41	2.17
17	3.10	2.52	2.27
18	3.23	2.69	2.36
19	3.40	2.83	2.45
20	3.56	2.97	2.54

* From "Population Movements in Sweden in Recent Years," by Carl-Erik Quensel, *Population Studies*, June 1947, p. 35.

Such tables have the advantage of taking into account the past fertility history of women in the cohort by duration of marriage. Calculations based on such tables broaden the base period from the fertility experiences of the most recently recorded year to that of the whole range of fertility history of married women in the childbearing ages. By making this shift in emphasis, such rates have a distinct advantage in that they predict the extension of a much longer and more stable experience than that of a particular year. Also, they predict the extension of an experience in which the influence of previous births to the mother is taken into account. If the principal determinator of long-time reproduction is the average number of children which married

couples will rear during the whole span of their life, then the marriage rates based upon the number of children ever born by duration of marriage begin to develop a body of data which can provide the answer. However, the question of predicting fertility is broader than that of predicting married fertility alone. For general fertility, the answer must take into account variation from time to time in the proportions of married people at various ages.

Unfortunately, also, the dates at which duration of marriage-fertility data have become available are so recent, even in most European countries, that insufficient time has elapsed for a reliable trend to appear. In the United States, reliable fertility data by duration of marriage are not available in any form. Year-to-year cross sections could be obtained from Census enumerations if questions were added as to the duration of marriage and these were cross-tabulated with the items of age and number of children ever born. Addition of the question of duration of marriage to the birth certificate in this country would not by itself solve the problem, for it would have to be supplemented by accurate nuptiality tables. Otherwise, there would be no population base upon which the rate could be calculated, and technical difficulties would be injected by reason of plural marriages and broken marriages.

Besides the short span of available data, there are other technical complexities in constructing marriage-adjusted fertility rates. Not the least of these is the problem of handling illegitimate births. This difficulty is usually recognized by calculation of separate rates for legitimate and illegitimate fertility. Another difficulty arises from the dissolution of marriages. If, as indicated above, the ex-married are progressively eliminated from the nuptiality table, a bias is probably introduced, for the reason that the complete fertility of married couples who remain married is greater than that of couples whose married life is interrupted during the childbearing period by death, separation, or divorce. The Census approach offers a possibility of minimizing this difficulty by the combination of children ever born to the ex-married with children born to the married.

In the United States, the fact that Census fertility data are secured only from married and ex-married women and the fact that young children are not completely enumerated reduces the accuracy of such information from enumerations. The addition of the duration of marriage question to the birth certificate would run counter to the established social policy of eliminating facts as to illegitimacy from birth records.

Theoretically, the objection to the use of number of children ever born by duration of marriage (unless combined in some fashion with the number of illegitimate births) arises from the fact that it ignores the changes in the percentage of the population who are married, because the base of the rate is narrowed from all women to married women. The rates do, however, have the advantage of getting away from a single year's base and the additional advantages of taking into account previous fertility history and providing for analysis of the effect of duration of marriage as a separate factor.

Since post-war European investigations have relied so extensively on marriage-adjusted rates, it behooves American students to follow this development carefully to determine whether it is not desirable to develop nuptiality tables and to secure reliable data by age, birth order, and duration of marriage. Whelpton, in his recent work, has corrected hypothetical cohorts for spinsterhood, but this correction has consisted of an arbitrary reduction of 10 per cent in the married population in each age, on the ground that this is the usual proportion of those who remain unmarried until the end of the childbearing period.

IV. GENERATION REPRODUCTION RATES[4]

The assumption underlying the net reproduction rate that the fertility and mortality rates of a single calendar year will remain invariable may be obviated by accumulating the actual number of female children born in each of the years of life of a generation; i.e., the generation gross rate is the sum of the age-specific gross female fertility rates obtained at dates which are advanced one calendar year for each advance of one year in the age of the mother. Similarly, appropriate generation mortality rates may be applied to determine the number of these daughters who will survive to the age at which their mother gave them birth. Illustration of the arrangement of rates and calculation is shown in Table 3. If a reliable Census enumeration of the number of children ever born to women who have lived through the childbearing period by various dates is available, a similar but somewhat cruder measure may be calculated from the number of births reported by women above age 45. This measure corresponds fairly well to the measure calculated from actual generation reproduction frequencies, except that in the United States, the Census enumerations tend to

[4] Depoid, M., "Reproduction Nette en Europe Depuis l'Origine de l'Etat Civil," *Etudes Demographique*, No. 1, Statistique General de la France.
Woofter, T. J., "Completed Generation Reproduction Rates," *Human Biology*, Vol. 19, No. 3, September 1947.

underestimate the number of children ever born, either because of faulty memory of respondents, or because of the tendency not to report illegitimate children or children of a previous marriage.[5] Such generation measures avoid the assumption of invariable fertility and mortality rates by substituting for the rates of a single calendar year the actual age-specific fertility rates of a particular cohort of women who live from ages 15 to 45. Another advantage is that they are based

TABLE 3

CHILDREN PER 1,000 WOMEN, GENERATION BEGINNING REPRODUCTION IN 1915

Age	Calendar Years	Children Per 1,000 Women				
		1st Year	2nd Year	3rd Year	4th Year	5th Year
16–19	1915–19	58	57	56	57	52
20–24	1920–24	166	166	153	152	154
25–29	1925–29	148	142	139	131	127
30–34	1930–34	95	89	85	80	82
35–39	1935–39	51	48	46	46	45
40–44	1940–44	17	17	15	16	17

Total births per 1,000 women living to age 45	2,507
Female births per 100 women	121.6
Female survival rate to age 28, 1933 life table*	.898
Net reproduction rate	1.092

* Instead of calculating separately the survival of daughters born to mothers at each age, accurate results may be obtained by calculating the survival of all daughters up to age 28 (the average age of mother) by a survival rate appropriate to the calendar year 18 years after the generation began reproduction. (Cf. "Generation Reproduction Rates," supra.)

upon the whole universe of women without elimination of the variations caused by percentages who are married. That is to say, they measure the impact of all factors operating upon fertility during a single generation of women because they are only standardized for age and mortality. They also have the advantage of allowing for the effect of past fertility performance on the fertility rates of the moment. Consequently, the generation rate is a slowly fluctuating measure in contrast with the rapidly changing rate of the calendar-year net reproduction rate. It, likewise, allows for the changes in the order of birth of children by following one group of women all the way through the childbearing period. In fact, the author has pointed out in the article previously cited that the generation method may be applied to the birth rates of children of the first, second, third, and higher orders, as well as to the total fertility rate. In these cases, the sum of the birth

[5] For such comparisons, see "Completed Generation Reproduction Rates," cited in Footnote 4.

rates of children of all orders in each age is equal to the gross reproduction rate.

An allowance for improvement in mortality by use of a generation life table results in a marked increase in net rates over those calculated on the assumption of the invariable continuation of current survival rates. The difference may be illustrated by the effect on the women born in 1900. When these women were 15 to 19 in the years 1915 to 1919, their children had a probability of surviving to age 17½ of only .878. The children born to the same generation of women when they were age 40 to 45 have a probability of surviving to age 17½ of .948,[6] an improvement of 8 per cent in 25 years (Table 4). In fact, the children born in 1940 had a probability of surviving to age 40 which is superior to the probability that children born in 1915 would survivive to age 5.

TABLE 4

5-YEAR GENERATION SURVIVAL RATES (MA) OF WHITE FEMALE CHILDREN UP TO AGE OF MOTHERS FOR GENERATIONS BEGINNING 1915–1940 AND 1920–1940

Mother's Age When Child Was Born	Initial Reproduction Year of Generation					
	1915	1920	1925	1930	1935	1940
15–19	.878	.900	.915	.927	.938	.948
20–24	.895	.910	.922	.934	.943	
25–29	.905	.917	.929	.938		
30–34	.911	.922	.933			
35–39	.917	.925				
40–44	.916					

It has been pointed out, however, that these rates have one serious disadvantage, namely, the longer time span covered. The complete generation experience of a cohort of women can only be recorded after age 45, which means that the child bearing experience extends back for 30 years. Likewise, the children who are born to women age 45 remain at risk of death for another 45 years, making the complete span of generation replacement performance in the neighborhood of 75 years.[7] This necessitates either carrying birth rates far into the past or

[6] In making such calculations for recent generations, it is necessary to project survival rates into the future to some extent. The generation life tables used in this and the previous article by the author on this subject were calculated by using actual survival rates up to 1940 and Whelpton's medium mortality assumptions thereafter.

[7] The average span is, however, much shorter. The author has pointed out in a previous article (cf. Footnote 4) that an accurate method of calculating generation mortality of the children born to a generation of mothers is to apply the single survival rate to age 28 (the average age at which mothers bore children) from a life table applying to a date 18 years after the generation begins child-bearing; i.e., the children born to women beginning childbearing in 1915 would survive at an average rate equal to survival to age 28 in 1933. Thus survival rates do not have to be estimated after the terminal childbearing age.

projecting them far into the future, and of making similar projections of survival rates. The whole United States was included in the birth registration area only in the 1930's. Whelpton has estimated rates back to 1920, and similar estimates for the group of mothers 15 to 19 may be made with some confidence back to 1915, with the result that only the generations beginning in 1915, 1916, 1917, 1918, and 1919 can be satisfactorily estimated. (These generations reached the age 44 in the years 1944, 1945, 1946, 1947, and 1948, and the daughters of the older mothers of these generations will be in their older childbearing ages from the years 1989 to 1993; but the average mortality of daughters of these generations may be closely approximated from life tables of 1933 to 1937.) It may be remarked in passing, however, that the same difficulty arises in converting gross rates based on the total number of children born by duration of marriage to net rates, for the reason that by the time a cohort of married women nears the end of the childbearing period, the duration of their marriage has been from 20 to 25 years.

V. COHORT REPLACEMENT RATES

The calendar year net reproduction rate is criticized on the ground that it projects temporary conditions too far into the future; the generation rate is criticized because it reflects conditions of a number of years past. The objections to both of these rates are partially obviated if a measure is used which reflects the cumulated number of births per 100 women of all childbearing ages up to a particular date, regardless of what age has been attained by that date. That is to say, in 1945 the women born in 1900 had attained age 45; those born in 1901 had attained the age 44; etc. Consequently, only the oldest cohort has completed childbearing; the others have varying periods to complete. If the cumulated reproduction of all reproducing cohorts[8] is set in relationship to a "standard" performance which would result in a net reproduction rate of 100, the result is an approximate measure of the rate of reproduction of all women of all ages up to the date at which the calculation is made.

As the cohort quota replacement rate has not previously been discussed in print, the following notation is introduced in relation to Tables 4, 5, and 6:

P_a = Female generation births per 100 women at age a.

M_a = Female generation survival rate up to age a.

[8] In the calculations presented in Tables 5 and 6, generations 5 years apart are shown instead of those beginning in every calendar year.

M_aP_a = Net reproduction at age a.

n = Number of generations in which an age group reproduces during the period for which the rate is calculated.

$$\overline{M_aP_a} = \frac{\sum M_aP_a}{n} = \text{Mean age-specific net reproduction.}$$

$$(1) \quad \sum_{15}^{45} \overline{M_aP_a} = \text{Average cohort replacement for all cohorts.}$$

$$M_aP_a{}^0 = \frac{\overline{M_aP_a}}{\sum_{15}^{45} \overline{M_aP_a}} = \text{Age replacement quota.}$$

$$(2) \quad \sum_{15}^{a+5} M_aP_a{}^0 = \text{Cohort replacement quota.}$$

The data for these calculations are arranged exactly as they are in the calculation of the generation net reproduction rate—i.e., by obtaining the female age-specific fertility rate for women of each age in the calendar year when they attained that age (Table 5) and multiply-

TABLE 5

FIVE-YEAR FEMALE BIRTHS PER 100 WOMEN IN EACH AGE—COHORTS BORN 1900–1925* (PA)

Age of Women When Children Were Born	Women Born in:					
	1900	1905	1910	1915	1920	1925
	Reached Age 15 in:					
	1915	1920	1925	1930	1935	1940
15–19	13.58	13.58	13.00	11.06	10.91	12.03
20–24	38.63	34.29	30.56	30.75	35.94	
25–29	33.42	29.10	28.28	33.66		
30–34	20.80	19.11	22.07			
35–39	11.45	11.83				
40–44	3.98					

* Based on revised estimates of P. K. Whelpton for under-registration and incompleteness of registration area. U. S. Bureau of the Census, "Forecasts of the Population of the United States," p. 17.

ing by the generation age-specific survival rate from 0 up to that age (Table 4). The results of these calculations are shown in Table 6 under the heading, "Surviving Female Children per 100."

With this arrangement of the data, the next step is to compute the average reproduction in each age ($\overline{M_aP_a}$). These averages are shown

TABLE 6

FIVE-YEAR NATIVE WHITE NET REPRODUCTION STANDARD QUOTAS AND NET REPRODUCTION FREQUENCIES UP TO 1944, FOR COHORTS BEGINNING REPRODUCTION IN 1915–1940

Age of Mother When Child Was Born (a to $a+5$)	Surviving Female Children per 100 (M_aP_a)						Mean Net Reproduction	Age Group Standard Quota
	Women Born in:							
	1900	1905	1910	1915	1920	1925		
	Reached Age 15 in:						$(\overline{M_aP_a})$	$(M_aP_a{}^0)$
	1915	1920	1925	1930	1935	1940		
15–19	11.92	12.22	11.89	10.24	10.11	11.40	11.30	10.79
20–24	34.33	31.20	28.18	28.72	33.89		31.26	29.85
25–29	30.25	26.68	26.29	31.57			28.70	27.41
30–34	18.95	17.61	20.57				19.04	18.18
35–39	10.59	10.94					10.72	10.24
40–44	3.70						3.70	3.53
Observed Cohort Reproduction $\sum_{15}^{a+5} M_aP_a$	109.65	98.65	86.93	70.53	44.00	11.40	104.72	
Cohort Standard Quota $\sum_{15}^{a+5} M_aP_a$	100.00	96.47	86.23	68.05	40.64	10.79	—	100.00
Cohort Replacement $\sum_{15}^{a+5} \frac{M_aP_a{}^0}{M_aP_a{}^0}$	1.096	1.023	1.008	1.036	1.083	1.057	1.047	

in the next to the last column of Table 6. The sum of these averages (Formula 1) yields an average net reproduction rate for all women of all childbearing ages from the time when they were 15 until the date of the calculation.

This calculation may be extended to determine a quota of "normal" reproduction to which each individual cohort may be compared in order to measure its relative performance up to the date of the calculation. These quotas are established as follows: Divide each of the average age-specific reproduction rates ($\overline{M_a P_a}$) by the total of these rates, thus converting the rates for each age into the percentage of births which normally occur in that age. Since the sum of the percentages equals 1, they determine age-specific frequencies which, if equaled, would result in a net reproduction rate of 1; thus, the extent of deviation of actual reproduction from such a quota measures the extent of the deviation of reproduction from a standard stationary rate. These quotas are shown in the last column of Table 6, and are cumulated in the next to the last line of Table 6 for comparison with actual cumulated reproduction. It is thus possible to compare any specific age in any cohort with a corresponding quota, or to compare the cumulated reproduction of each cohort with its quota.

Whereas the major weighting of the experience of a completed generation is 17 years before the end of their experience (when women are age 28), the major weighting of the experience of the total population is about 11 years preceding the date of the calculation. It will be noted that in Table 6, 11 of the 21 cohort-age groups measure fertility for the 10 years immediately preceding the date of the calculation, and 10 of the 21 relate to the older generations whose experience extends from 10 to 30 years back of the date of calculation. A series of such cohort replacement rates is even more stable than the rate for a series of generations.

In avoiding some of the disadvantages of other rates, however, such calculations also lose some of their advantages, since they refer neither wholly to the present nor wholly to the complete experience of generations. However, a series of such calculations extending over a number of years would provide a basis for extrapolation which is more sensitive to present conditions than a series of generation rates.

Whelpton presented fertility data for such incomplete cohorts in a paper as yet unpublished.[9] He did not, however, convert the gross reproduction rates of these incomplete cohorts into net rates.

Paper delivered at the 1947 Session of the International Statistical Institute.

VI. PARITY ADJUSTED RATES[10]

Although Whelpton has published rates adjusted for age, parity, fecundity, and marriage, we shall discuss at this point only the adjustments which were made for parity. His observations led to the conclusion that, if birth rates by order of birth, as calculated for a single abnormal calendar year, are assumed to apply to a cohort of women who actually pass through the childbearing period, impossible results can be obtained. Such calculation for the year 1942 produced the impossible rate of 1,084 first births per 1,000 women. (With the 1947 calendar year as a basis, the results would be still more impossible.)

His first approach to the solution of this difficulty was to standardize for birth order by beginning with the life table cohort of women who had had no children and successively deducting the mothers who bore a first child and survived to the next age from the base of the rate for second births, and so on for each higher order of birth. The resulting parity-adjusted rate had the advantage of avoiding impossible results, but was still open to the objections which have been made to rates which generalize the experience of a single calendar year. Realizing this, Whelpton in some of his later (unpublished) work has made calculations according to order of birth on the basis of the actual recorded generation experience. If recorded statistics are available, this method has all of the advantages and disadvantages of the generation ratio described in the preceding section. The period for which Whelpton has made his later generation birth order calculations is the same as that used for the generation reproduction rates calculated by the author. They are thus susceptible to reduction to net rates by the use of generation life tables and have the obvious advantage of being adapted to studies in which birth order is an important factor singled out for special consideration.

VII. CONCLUSION

The variety of experimentation with rates in all countries where data are available should provide the basis for extending and refining the analysis of fertility. The development of a variety of techniques should, therefore, be welcomed. A review of the foregoing discussion reveals that no one of these rates is most appropriate for all purposes, but that each is well adapted to some particular purpose. Perhaps the effort of some demographers to arrive at a single "optimum" measure which will reflect the condition of the moment and at the same time

[10] Whelpton, P. K., "Reproduction Rates Adjusted for Age, Parity, Fecundity, and Marriage," *Journal of the American Statistical Association*, December 1946.

provide a basis for the measurement of underlying fertility trends is as difficult as it is "to eat one's cake and have it, too." Certainly no measure reviewed in this article is well adapted to both purposes. Demographers who are interested in "invariable fertility trends" might do well to review the history of economic prediction which relies not on one but on the variety of interacting indices.

It will be observed that the rates summarized above fall into two general groups:

(1) Those which use as their basis the experience of a single calendar year and thus concentrate on the condition of the moment. They are subject to violent fluctuations, such as those that have taken place in the last 20 years. These fluctuations make it difficult to select the proper base period from which to project a trend and to determine the shape of the curve which is being projected.

(2) Those which follow the generation technique of recording the actual fertility experience of mothers at a particular age in the year during which they attained that age. These rates include the general generation rate, the rates based on the cumulated number of children born to women who were married at a certain date and age, and the cumulated birth order rates based on actual chronological experience. This latter type is not so sensitive to year-to-year changes, and, consequently, does not reflect conditions of the immediate present as accurately as do rates based on the conventional net reproduction technique. They are, however, more stable and are based on a longer time period, reflecting the impact of past fertility on present rates. Not the least of the advantages of the calculations of the generation type is the fact that they are adapted to conversion to net rates by allowing for generation improvement in mortality.

ON ESTIMATING THE MEAN AND STANDARD DEVIATION OF TRUNCATED NORMAL DISTRIBUTIONS*

A. C. Cohen, Jr.
University of Georgia

The problem considered is that of estimating the mean and standard deviation of a normally distributed population from a truncated sample when neither count nor measurements of variates in the omitted portion of the sample is known. Formulas are developed whereby certain special functions required in solutions given for this problem by Karl Pearson and Alice Lee and by R. A. Fisher may be readily evaluated with the aid of an ordinary table of the areas and ordinates of the normal curve. A method of successive approximations is illustrated which, with the aid of the above formulas permits the utilization of either the Pearson-Lee or the equivalent Fisher method to obtain the desired estimates with an improvement in accuracy regardless of whether or not the special tables ordinarily required by these two methods are available.

I. PEARSON-LEE METHOD

Karl Pearson and Alice Lee [1], and Alice Lee [2] employed the method of moments as early as 1908 to develop formulas which may be used to estimate the mean and standard deviation of a normally distributed population from data provided by a truncated random sample from which all record including both count and measurements of all variates whose value is beyond a given truncation point, has been omitted. Their results except for minor changes in notation may be summarized as follows:

$$m' = x_0' - h'\sigma' \tag{1}$$

$$\sigma' = \frac{1}{n}\sum x\cdot\psi_2 \tag{2}$$

$$\frac{n\sum x^2 - (\sum x)^2}{(\sum x)^2} = \psi_1. \tag{3}$$

In the above equations, m' and σ' are estimates of the mean and standard deviation of the population (complete distribution). x_0' is the point of truncation measured on the original scale of the variate x'. The omitted portion of the sample is here considered to be to the left of x_0'. The summations $\sum x$ and $\sum x^2$ are taken about x_0' as an origin.

* A portion of this paper was presented before the Southeastern Section of the Mathematical Association of America at Tuscaloosa, Alabama, March 19, 1949.

h' is the point of truncation measured in standard units of the population; that is,

$$h' = \frac{x_0' - m'}{\sigma'}.$$

n is the number of variates in the truncated sample. ψ_1 and ψ_2 are moment functions of h'. Tables of both these functions evaluated to three places of decimal at intervals of 0.1 in h' are contained in the original papers and also in "Tables for Statisticians and Biometricians" [3] Vol. I. Table XI and Vol. II. Table XII. To apply the Pearson-Lee method, one proceeds as follows:

(1) Evaluate the left side of Equation (3) from the sample data. Enter the table of ψ_1 with this value and obtain h' by inverse interpolation.

(2) Using the above value of h' as the argument, read ψ_2 from the appropriate table and apply Equation (2) to obtain σ'.

(3) With both σ' and h' determined, then apply Equation (1) to obtain m'.

Unfortunately the tables required for this method are not as widely distributed as might be desired. Furthermore the entries contain too few significant digits and are tabulated at an interval of the argument (0.1) that is too wide to permit h' to be determined with sufficient accuracy for many applications.

II. FISHER METHOD

In 1931, R. A. Fisher [4] demonstrated that the "Maximum Likelihood" estimates for this problem are identical with those obtained by the method of moments. His results, however, were expressed in a slightly different form from those of Pearson and Lee. Fisher employed a moment function of h' (designated by ξ in his discussion) which he labeled as an I_n function and which may be defined as:

$$I_n(h') = \frac{1}{\sqrt{2\pi}} \int_{h'}^{\infty} \frac{(t - h')^n}{n!} e^{-t^2/2} dt \tag{4}$$

and for which the following relations hold:

$$(n + 1)I_{n+1} + h'I_n - I_{n-1} = 0; \qquad n > -1. \tag{5}$$

$$\frac{dI_n}{dh'} = - I_{n-1}. \tag{6}$$

The Fisher results in terms of the I_n functions are:

$$\sigma' = \frac{1}{n} \sum x \frac{I_0}{I_1} \tag{7}$$

and

$$\frac{n \sum x^2}{(\sum x)^2} = \frac{2I_0 I_2}{I_1^2} \tag{8}$$

His equation for obtaining m' after σ' and h' are determined is the same as that given by Pearson and Lee (Equation 1.)

Tables of I_0, I_1 and $I_0 I_2/I_1^2$ (labeled $Hh_0Hh_2/(Hh_1)^2$) as required for use in the Fisher formulas (Equations 7 and 8) are included in "Mathematical Tables" Vol. 1 of the British Association for the Advancement of Science. Entries in these tables are given to a greater number of decimals than the Pearson-Lee tables (from 6 to 9 significant digits for most entries) but the interval of the argument h' (0.1) is the same as for the Pearson-Lee tables. The greater number of significant digits permits greater accuracy in determining h', but only by resorting to inverse interpolation formulas involving the second and higher order differences. At best such computations are rather tedious and somewhat bothersome to carry out. With regard to availability, the B.A.A.S. Tables are perhaps even less widely distributed than the Pearson Tables. The application of the Fisher results is almost identical with that of the Pearson-Lee results. The quantity on the left side of (8) is computed from the sample data and h' is determined by inverse interpolation from the table of $I_0 I_2/I_1^2$. Using the value of h' thus determined, I_0 and I_1 are obtained from the appropriate tables and σ' is computed by use of Equation (7). Equation (1) is then used as before to obtain m'.

III. EQUIVALENCE OF PEARSON-LEE AND FISHER RESULTS

Since Equations (2) and (3) are equivalent to (7) and (8) it follows that:

$$\psi_2 = I_0/I_1 \tag{9}$$

and

$$\psi_1 = \frac{2I_0 I_2}{I_1^2} - 1. \tag{10}$$

IV. NEW CONTRIBUTIONS

In the present paper, equations are derived which permit the calculation of ψ_1, ψ_2 and likewise $2I_0 I_2/I_1^2$ without resort to any tables other than an ordinary table of areas and ordinates of the normal frequency curve such as can be found in practically any handbook of

mathematical tables. By using the formulas presented herein, the Pearson-Lee or the Fisher technique can be readily applied regardless of the availability of the special tables previously mentioned. Even when the special tables are available, the formulas developed in this paper permit the attainment of greater accuracy in determining h' and consequently in obtaining σ' and m' with a minimum computing effort.

V. DERIVATIONS

Let $n=0$ and 1 respectively in Equation (5) to obtain

$$I_1 = I_{-1} - h'I_0 \tag{11}$$

and

$$2I_2 = I_0 - h'I_1. \tag{12}$$

Let $n=0$ in Equation (6) and we have $dI_0/dh' = -I_{-1}$. If now $n=0$ in Equation (4) it follows that

$$I_0 = \frac{1}{\sqrt{2\pi}}\int_{h'}^{\infty} e^{-t^2/2}dt \tag{13}$$

which is recognized as the area under the normal curve to the right of the ordinate $t=h'$. Direct differentiation of (13) gives $dI_0/dh' = -\phi(h')$ where $\phi(h')$ is the ordinate of the normal curve at $t=h'$.

$$\left(\phi(h') = \frac{1}{\sqrt{2\pi}}e^{-h'^2/2}\right).$$

Consequently we may write

$$I_{-1}(h') = \phi(h'). \tag{14}$$

Upon substituting the results of Equations (11), (12) and (14) in the right side of Equation (8) it follows that

$$\frac{2I_0I_2}{I_1^2} = \frac{[I_0 - h'(\phi - h'I_0)]I_0}{[\phi - h'I_0]^2}. \tag{15}$$

Now if we define

$$Z(h') = \frac{\phi(h')}{I_0(h')}, \tag{16}$$

and divide both the numerator and denominator of the right side of (15) by I_0^2 we obtain

$$\frac{2I_0I_2}{I_1^2} = \left(\frac{1}{Z - h'}\right)\left(\frac{1}{Z - h'} - h'\right). \tag{17}$$

From Equation (10) it then follows that

$$\psi_1 = \left(\frac{1}{Z - h'}\right)\left(\frac{1}{Z - h'} - Z\right). \tag{18}$$

Similarly from Equations (11) and (14) we obtain

$$\frac{I_0}{I_1} = \psi_2 \quad \frac{1}{Z - h'} \tag{19}$$

Equations (17), (18) and (19) are thus expressed in convenient forms to permit rapid calculation of the special functions required by both the Pearson-Lee and the Fisher methods for any values of the argument h' as may be necessary, provided only that existing tables of the areas and ordinates of the normal frequency curve are available. These formulas would also be useful in extending both the Pearson-Lee and the Fisher tables by computing additional entries at closer intervals of the argument h'.

VI. DETERMINING h' BY SUCCESSIVE APPROXIMATIONS

Although tables are used in both the Pearson-Lee and the Fisher methods, the basic problem involved is the solution of either Equation (3) or its equivalent Equation (8) for h'. A method of successive approximation which makes use of (17), (18), and (19) and simple linear interpolation has been found to be quite satisfactory for determining h' to additional significant digits. If either the Pearson-Lee or the Fisher tables are available they might be used to furnish a first approximation of h'. For use when neither of these tables are available, a graph of ψ_1 for values of h' and from -3.5 to $+3.5$ is given in Figure 1. In the absence of both tables and graph a reasonably satisfactory first approximation to h' can be obtained from

$$h' \sim (x_0' - \bar{x})/s_x \tag{20}$$

where $\bar{x}$ is the mean and s_x is the standard deviation of the truncated sample. To compute s_x use the formula

$$(ns_x)^2 = n\sum x^2 - (\sum x)^2.$$

VII. NUMERICAL EXAMPLE

The various steps involved in determining h' and subsequently σ' and m' by the successive approximation technique mentioned in the previous paragraph can best be understood by computing these quantities for a typical set of data such as the following:

$$n=37;\ x_0'=0.850; \sum x=51.8600; \text{ and } \sum x^2=98.0156$$

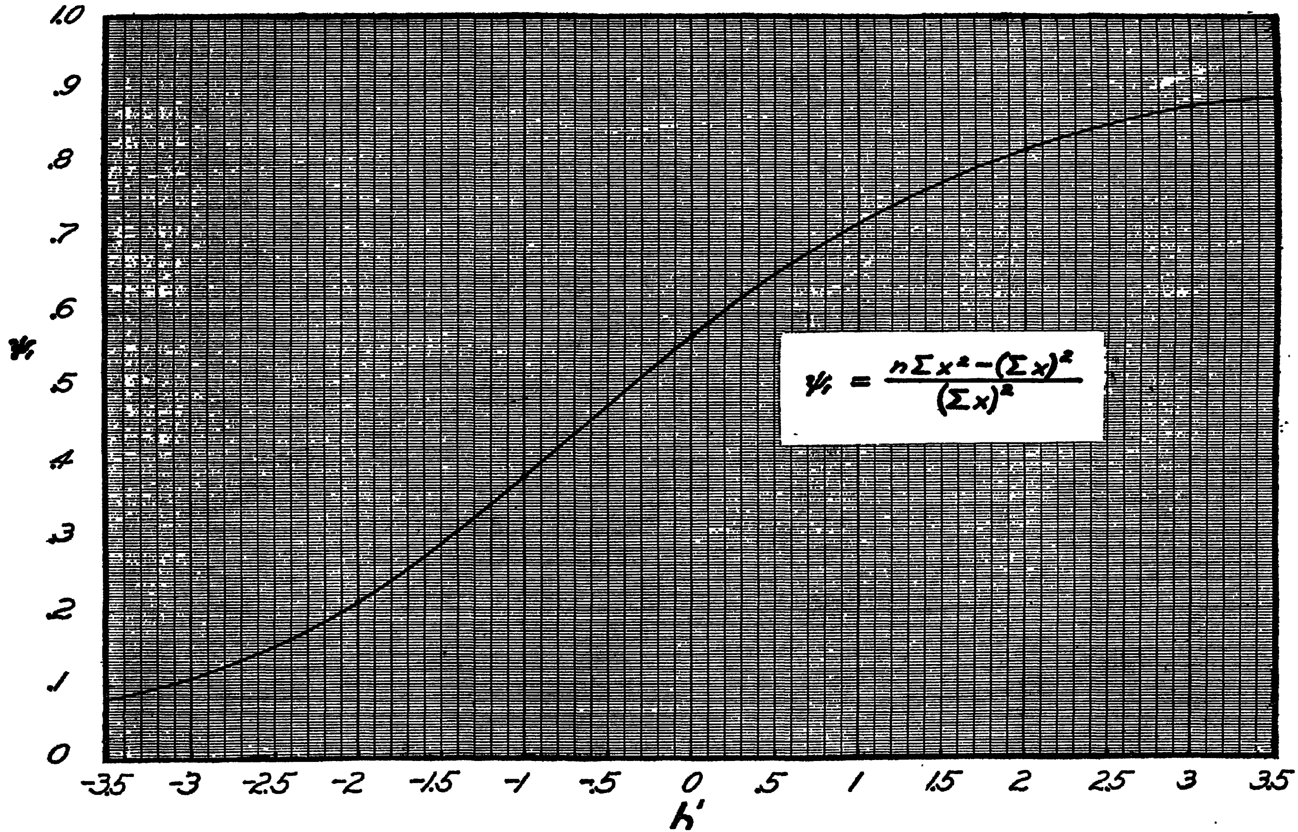


FIGURE 1

GRAPH SHOWING RELATION BETWEEN ψ_1 AND h'

where x has been measured from the terminus as an origin ($x_i = x_i' - x_0'$). Thus $\bar{x} = 1.401622$; $s_x = 0.83$; and

$$\frac{n\sum x^2 - (\sum x)^2}{(\sum x)^2} = 0.348441.$$

From the graph of Figure 1, we find $h_1' = -1.20$ as a first approximation. By direct substitution in Equation (18) it is found that $\psi_1(-1.20) = 0.341734$. Since this value is less than 0.348441 and since ψ_1 is an increasing function of h', it is necessary to select a value greater than -1.20 as a next approximation. We then find that $-1.20 < h' < -1.10$. Linear interpolation gives $h_2' = -1.16$ as a closer approximation and following the next step we establish that $-1.165 < h' < -1.164$. Again linear interpolation gives $h_3' = 1.1643$ which is accepted as a final approximation since this determination is sufficiently accurate for our purposes.

The tables of areas and ordinates of the normal distribution used in making these computations contained six significant digits and were tabulated at intervals of 0.01 for the argument. In using these tables it was necessary to employ interpolation for obtaining entries only for the two final approximations. Table 1 details the computations involved in the various steps described above.

TABLE 1

h'	I_0	ϕ	Z (3) ÷ (2)	$Z-h'$ (4) − (1)	ψ_2 1/(5)	$\frac{1}{Z-h'} - Z$ (6) − (4)	ψ_1 (6)×(7)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
−1.10	.864334	.217852	.252046	1.352046	.739620	.487574	.360619
−1.20	.884930	.194186	.219436	1.419436	.704505	.485069	.341734
−1.16	.876976	.203571	.232128	1.392128	.718325	.486197	.349247
−1.17	.879000	.201214	.228912	1.398912	.714841	.485929	.347362
−1.164	.877786	.202628	.230840	1.394840	.716928	.486088	.348490
−1.165	.877988	.202392	.230518	1.395518	.716580	.486062	.348302

It has been the writer's experience that by systematically arranging computations as shown in the above table the tedium of making each individual determination of $\psi_1(h')$ is considerably reduced.

By interpolation from Table 1 it is readily found that

$$(Z - h')\big|_{h'=-1.1643} = 1.395043$$

and it follows from Equations (2) and (19) that

$$\sigma' = 1.401622/1.395043 = 1.0047.$$

From Equation (1) it then follows that $m' = 0.850 - (1.0047)(-1.1643) = 2.020$. This completes the solution.

VIII. VARIANCE OF ESTIMATES

R. A. Fisher obtained the following formulas for the variance of σ and h' (Fisher's ξ)

$$V(\sigma) = \frac{\sigma^4\mu_2}{n\{\mu_2'\mu_2 + \sigma^2(2\mu_2 - \mu_2')\}}, \tag{21}$$

$$V(h') = \frac{\sigma^2(\mu_2' + \sigma^2)}{n\{\mu_2'\mu_2 + \sigma^2(2\mu_2 - \mu_2')\}}, \tag{22}$$

and for the correlation between the sampling errors of σ and h'

$$r_{\sigma,h'} = \frac{+\sigma\mu_1'}{\sqrt{\mu_2(\mu_2' + \sigma^2)}}. \tag{23}$$

where μ_κ' is the kth moment of the truncated sample about its terminus x_0' and μ_κ is the kth central moment of the truncated sample.

It can be shown that in terms of the function Z defined by Equation (16) that the above equations become

$$V(\sigma) = \frac{\sigma^2}{n}\left[\frac{1 - Z(Z - h')}{[1 - Z(Z - h')][2 - h'(Z - h')] - (Z - h')^2}\right], \tag{24}$$

$$V(h') = \frac{1}{n}\left[\frac{2 - h'(Z - h')}{[1 - Z(Z - h')][2 - h'(Z - h')] - (Z - h')^2}\right], \tag{25}$$

$$r_{\sigma,h'} = \frac{+(Z - h')}{\sqrt{[1 - Z(Z - h')][2 - h'(Z - h')]}}. \tag{26}$$

REFERENCES

[1] Pearson, Karl and Lee, Alice, "On the Generalized Probable Error in Multiple Normal Correlation," *Biometrika*, Vol. 6, 1908, pp. 59–68.

[2] Lee, Alice, "Table of Gaussian 'Tail' Functions When the 'Tail' is Larger Than the Body," *Biometrika*, Vol. 10, 1915, pp. 208–215.

[3] Pearson, Karl, editor (1914, 1931). *Tables for Statisticians and Biometricians*, Part I (1914, 3rd edition 1930) and Part II (1931), Cambridge University Press.

[4] Fisher, R. A., "Properties and Applications of *Hh* Functions," Introduction to *Mathematical Tables*, Vol. 1, British Assn. for the Advancement of Science, 1931, pp. xxvi–xxxv.

ON SOME MATHEMATICAL PROBLEMS ARISING IN THE DEVELOPMENT OF MENDELIAN GENETICS*

HILDA GEIRINGER
Wheaton College, Massachusetts

In this paper some basic problems of theoretical Mendelian genetics are presented. The general situation of an arbitrary number of linked loci for "diploids" (normal organisms) as well as for "autopolyploids" is considered, under random mating. Particular stress is laid upon formulating the problems and results within the framework of probability theory, expressing them in terms of three basic distributions, the distributions of genotypes and of gametes, and the segregation distribution. The task of the mathematical theory consists in establishing recurrence relations for the distributions of gametes and of genotypes, in integrating them and in investigating the limit behaviour of those distribution as n, the number of discrete non overlapping generations tends toward infinity.

IN A PAPER delivered in 1935 at the Massachusetts Institute of Technology, J. B. S. Haldane [9, g] spoke on "Some Problems of Mathematical Biology." In this lecture, which covers a much broader subject than the present paper, he suggests the following classification of the problems of mathematical biology. A first group of problems is concerned with the life of the cell; next, one may consider the analysis of a whole organism composed of cells, like a tree or an animal; on a third level, one studies the mutual relations of a number of organisms, —members of the same brood, individuals of the same population,— and investigates the biological fate of such populations.

The problems dealt with in the present paper belong to the third group, which is concerned with entire populations. The mathematical tools for these investigations are offered by probability theory and statistics, since this is the branch of mathematics which deals with mass phenomena. However, we will not consider here the manifold problems connected with the statistical evaluation of biological observations. From a systematic standpoint problems of biological statistics are not different from the statistical problems occurring in connection with other series of observations. The more specific task of mathematical genetics is comparable to what is done in the kinetic theory of gases, in the theory of quanta, etc. One starts by formulating probability laws that express certain biological facts and then tries to deduce

* Lecture delivered January 1948 at the University of Chicago.

in a mathematical way consequences which are to be checked by observations. The simplest example of such a procedure presents itself in Mendel's theory of the heredity of a single character.

1. RANDOM INHERITANCE OF ONE CHARACTER. THE BASIC PROBABILITY DISTRIBUTIONS

Mendel [17] studied certain single traits or characters, like the heredity of the color of the flower of peas or the size (tall or short) of this plant, by actually counting the numbers of each type of progeny which resulted from a given cross, thus applying statistics to the phenomena of inheritance. His fundamental discoveries were made on the garden pea, *Pisum sativum.* Systematic observations and analysis led him to the following daring hypothesis. The visible color (red or white in case of the peas,—red, white or pink in case of four-o'clocks) is dependent on a pair of factors which we may denote for the moment by R and W. Each fertilized zygote (and each cell of the organism into which it develops) contains *two* of these factors and may thus be of the type RR, or RW, or WW. The gametes (egg and sperm) contain one factor only, selected at random from among the two factors contained in the cell out of which the gamete is formed. A new zygote of the following (filial) generation is then formed by the random union of two gametes, and, consequently, again contains two factors, and so on.

Such factors have been called Mendelian factors, unit factors, or *genes.* There is no doubt, today, that the genes are located in the *chromosomes.* Two genes alternative to one another, like R and W, are spoken of as allelomorphic factors, or *alleles.* There may be more than two members of such an "allelomorphic series" e.g. t = tall, s = short, and d = dwarf with respect to size. A zygote which contains two alleles of the same kind, like RR, or WW, or tt, etc., is said to be *homozygous* for the factor in question while an RW-plant or a plant of type td, is called *heterozygous* or *hybrid. Locus* means the particular place in a particular chromosome at which there are alleles. *Character* relates to the effect of genes. A single character, like color, may be affected by many loci and, on the other hand, the same locus may influence several visible characters. In the case of the flower-color of peas, however, this character is determined by *one* locus.

In the first paragraph of this section we have presented the essential content of Mendel's famous *first law.* Let us attempt a more mathematical formulation. We assume that, corresponding to each locus, there exists a *random variable,* x, which may take on r distinct values,

the r alleles, $x=a_1, a_2, \cdots, a_r$. In the example of the color of peas, $r=2$; here a_1 stands for "red" and a_2 for "white." *Three* alleles determine the human blood groups, and there are at least $r=14$ known alleles for the eye color of *Drosophila melanogaster*, the small vinegar fly which plays as important a role in modern genetics as did Mendel's peas a century ago. With respect to such a locus the genetic type of an organism (zygote as well as grown organism) is specified, not by one, but by *two* values, x and y, of this random variable which may or may not be equal to each other. They represent the organism's maternal and paternal heritage, since it receives one allele from its mother and the other from its father.

We assume that two organisms are genotypically the same with respect to the locus in question if in one of them x comes from the mother and y from the father, while in the other organism it is the other way round. If we denote the type of an organism by (xy), where the first letter denotes the maternal and the second the paternal heritage, this assumption reads,

$$(xy) = (yx). \tag{1}$$

Consequently there are $r(r+1)/2$ possible genotypes in this case. If $r=2$ the three types are often denoted by (AA), (Aa), and (aa). Next we suppose *distinct, non-overlapping generations*. A basic assumption is then that in a certain "initial" generation the possible types (xy), $(x=a_1, \cdots, a_r)$; $(y=a_1, \cdots, a_r)$ are distributed according to a *law of probability*. This distribution will be called the initial probability *distribution of genotypes*, $w^{(0)}(xy)$. From it, by means of the hypotheses which characterize our problem, the distributions of genotypes for later generations will follow. We denote the distribution for the nth generation by $w^{(n)}(xy)$. In accordance with (1) we must assume that

$$w^{(n)}(xy) = w^{(n)}(yx) \qquad (n = 0, 1, \cdots), \left(\begin{matrix} x \\ y \end{matrix} = a_1, \cdots, a_r\right). \tag{2}$$

There is no loss of generality if we suppose that the initial distribution of genotypes is the same for males and females. In fact, it is easily seen that under random mating any difference between initial distributions disappears in the first filial generation. For this distribution we have

$$\sum_x \sum_y w^{(n)}(xy) = 1, \qquad (n = 0, 1, \cdots), \left(\begin{matrix} x \\ y \end{matrix} = a_1, \cdots, a_r\right). \tag{3}$$

Next Mendel assumed that in the formation of a new individual the

parent (through the gamete) transmits to the offspring one of its two genes, either x or y. The choice between the two values happens according to a probability law. Mendel's assumption, deduced from and confirmed by observation, is that *the two probabilities* for the segregation of either of the two genes *are equal.* With a view to more general cases, we introduce a second basic distribution, the *segregation distribution.* Denote by l_0 the probability that the paternal gene be transmitted, by l_1 the analogous probability for the maternal gene. Then, in the particular case under consideration

$$(4) \qquad l_0 + l_1 = 1, \qquad l_0 = l_1 = \tfrac{1}{2}.$$

The segregation distribution is not so trivial in more general situations. We shall however retain two of the assumptions made here: (a) The segregation distribution is independent of "n", it does not change through the generations. (b) It does not depend on the genotype of the parent. We shall see that the segregation distribution plays a basic role wherever random mating is considered. Certain problems where selection is involved cannot be described in terms of the segregation distribution. (See section 2.)

From these two distributions we derive the third important distribution, the *distribution of gametes,* $p^{(n)}(x)$, $(n=0, 1, \cdots)$, $(x=a_1, \cdots, a_r)$. This is the probability that the gamete be of constitution x, i.e. possesses the gene x. Let us compute this distribution. In order to transmit the gene x the parent must possess this gene and transmit it. Accordingly we have for example for the gene $x=1$, (writing $1, 2, \cdots, r$ instead of $a_1, a_2, \cdots, a_r$), $p^{(n)}(1)=w^{(n)}(11)(l_0+l_1)+w^{(n)}(12)l_1+w^{(n)}(21)l_0+ \cdots +w^{(n)}(1r)l_1+w^{(n)}(r1)l_0$. This formula is easily understood. For example, the term $w^{(n)}(12)l_1$ is the probability that the parent possesses the maternal gene "1" and the paternal gene "2," multiplied by the probability, l_1, that the transmitted gene be the maternal gene. Obviously the sum of all such terms, as contained on the right side of the preceding formula, gives the probability of a gamete of type "1," (if we assume certain circumstances which we shall analyze presently). Because of (2) and (4) the formula may be written as follows

$$(5) \qquad p^{(n)}(x) = \sum_{y}^{1\cdots r} w^{(n)}(xy) \qquad (x = 1, \cdots, r, \ n = 0, 1, \cdots)$$

with

$$(5') \qquad \sum^{1\cdots r} p^{(n)}(x) = 1.$$

A new individual of the $(n+1)$st generation is formed by the fusion of two gametes of the nth generation. The assumption that this fusion happens at random is expressed in the formula:

$$w^{(n+1)}(xy) = p^{(n)}(x)p^{(n)}(y) \qquad \left(\begin{matrix}x\\ y\end{matrix} = 1, 2, \cdots, r\right) \tag{6}$$

and that finishes the cycle, since we now know $w^{(n+1)}(xy)$, the distribution of genotypes in the next generation.

Before analyzing the assumptions which underlie these formulas I want to derive the famous *law of constancy of gametic proportions*, first recognized by the biologist W. Weinberg [25] and proved by the mathematician G. H. Hardy [10]. We obtain from (5) and (6):

$$\begin{aligned} p^{(n+1)}(x) &= \sum_{y}^{1\cdots r} w^{(n+1)}(xy) \\ &= \sum_{y}^{1\cdots r} p^{(n)}(x)p^{(n)}(y) = p^{(n)}(x) \sum_{y}^{1\cdots r} p^{(n)}(y) \\ &= p^{(n)}(x) \qquad\qquad (n = 0, 1, \cdots). \end{aligned} \tag{7}$$

From this it follows that

$$w^{(n+1)}(xy) = w^{(n)}(xy), \qquad (n = 1, 2, \cdots). \tag{7'}$$

These basic results state that *the distribution of gametes remains constant throughout, while the distribution of genotypes remains constant from the first filial generation on*. These results hold in this form in our "simplest case" only. In more complicated problems the first result must, in general, be modified and the second fails.

2. REMARKS ON "RANDOM MATING"

Let us try to give a definition, or rather a mathematical explanation, of the concept of random mating as opposed to "selection." I do not propose to discuss the meaning in probability of the concepts "random" or "randomness" or "random variable," which are inherent in any probability theory. Although there is much controversy with respect to these notions, every statistician attributes a certain meaning to them.

It seems to me that the mathematical meaning of "random mating" is contained in the equations (5) and (6) by which $w^{(n\neq 1)}$ is derived from $w^{(n)}$. In an equation of type (5), the distribution of gametes in the nth generation appears as a linear expression in the $w^{(n)}(xy)$, the values of the distribution of zygotes, with constant coefficients which

are sums of values of the segregation distribution. An equation of type (6) expresses the random fusion of the two gametes. In the various cases of selection and mutation *one or more of the assumptions which lead to* (5) *and* (6) *cannot be maintained.* We shall see, for example, that these equations contain the assumption that within the population there are no genotypic differences with respect to the attainment of maturity, to fecundity, or to mortality (no matter whether such differences are "natural" or "planned"). Another assumption contained in our equations is that the choice of the mate happens at random. Two examples will illustrate these points.

Consider, as in section 1, the "simplest case" of one locus and assume $r=2$ alleles. Also suppose that the choice of the mate is still due to chance, but that there is a differential viability for the three genotypes. How do we express this mathematically?

Write A and a for the two alleles and introduce for the sake of brevity

$$w^{(n)}(AA) = p_n, \qquad w^{(n)}(Aa) = q_n, \qquad w^{(n)}(aa) = r_n.$$

These p_n, q_n, r_n constitute the distribution of zygotes as generated by their parents. In case of a different viability of the types we can no longer state that the distribution of genotypes at the time of maturity is the same as at time of birth. We have to introduce a *different* distribution, p_n', q_n', r_n', the distribution of the parents of the next generation. Various types of selection as indicated above may be accounted for in this way. We define p_n', q_n', r_n' by means of "selection coefficients" α, β, γ (only the proportions of these three numbers matter) $p_n':2q_n':r_n'=\alpha p_n:2\beta q_n:\gamma r_n$, with $p_n'+2q_n'+r_n'=1$. The distribution of gametes is now deduced from the p_n', q_n', r_n', rather than from the p_n, q_n, r_n. Writing x_n and y_n instead of $p^{(n)}(A)$ and $p^{(n)}(a)$ the formula which takes the place of (5) is:

$$x_n = p_n' + q_n', \qquad y_n = q_n' + r_n'.$$

Although (5) no longer holds, the rule expressed in (6) is still valid:

$$p_{n+1} = x_n^2, \qquad q_{n+1} = x_n y_n, \text{ etc.}$$

Mathematically this is a problem of a different character, with essentially different results from those of the problem of section 1.

We quote a second example where the random choice of the mate no longer applies and differential viability or an equivalent condition is not assumed. The strongest deviation from the concept of random choice of the partner presents itself if the genotype of the mate is uniquely determined by the individual's own genotype. Assume that

there exists the rigorous "rule" (natural or artificial) that only identical types may mate. Then there are no longer six but only three possible types of matings: $AA\times AA$, $Aa\times Aa$, and $aa\times aa$. With our original notation we have:

$$w^{(n+1)}(AA) = w^{(n)}(AA) + \tfrac{1}{2}w^{(n)}(Aa),\ w^{(n+1)}(Aa) = \tfrac{1}{2}w^{(n)}(Aa),\ \cdots .$$

The third equation is similar to the first, with A and a interchanged. These equations replace (5) *and* (6). To understand the new situation we write for the right side of the first of these equations: $w^{(n)}(AA)\cdot 1\cdot 1 +2w^{(n)}(Aa)\cdot 1\cdot \frac{1}{4}$. Consider e.g. the second of these two terms. Here $2w^{(n)}(Aa)=w^{(n)}(Aa)+w^{(n)}(aA)$ is the probability of a female of type (Aa). To get the probability of the mating $Aa\times Aa$ we no longer multiply $2w^{(n)}(Aa)$ by $2w^{(n)}(Aa)$, as we would in random mating, but by "1," which represents the *conditional probability* that under the considered "law" the male partner be of type Aa if the female is of type Aa. Hence $2w^{(n)}(Aa)$ is the probability of the mating $Aa\times Aa$. Finally $\frac{1}{4}$ stands for the probability that the offspring of the mating $Aa\times Aa$ be of type AA. More generally, we may denote by $\pi_{\kappa\lambda}$ the probability of a union of the types κ and λ and by $\rho_{\kappa\lambda}{}^{\mu}$ the probability that the offspring of a mating (κ, λ) be of type μ. We assume that for all κ, λ: $\rho_{\kappa\lambda}{}^{\mu}=\rho_{\lambda\kappa}{}^{\mu}$ and $\sum_{\mu}\rho_{\kappa\lambda}{}^{\mu}=1$. Here $\rho_{\kappa\lambda}{}^{\mu}$ need not be, as in our example, the product of two segregation probabilities. It may depend on the types of both parents, whereas our segregation distribution refers to each parent separately and does not depend on the parental types (see sec. 4).

We do not attempt to present a scheme which would apply to the most general case of "selection." The preceding examples are intended to show how new situations call for new concepts.

3. SOME MATHEMATICAL PROBLEMS IN MENDELIAN GENETICS

We now consider some important generalizations of our "simplest problem." In this simplest case there exists no "recurrence problem," since after *one* good mixing the characteristic distributions do not change any more. This is not a typical result. However, the *concepts* introduced for this simplest case remain valid except for certain generalizations.

We continue to assume random mating. As an important example of a more general situation we first consider *various characters* of an individual simultaneously, as e.g. color of the flower, length of the stem, and shape of the seed, etc. The genetic constitution of the individual is still determined by the two gametes which represent its maternal

and paternal heritage. Each gamete, however, no longer consists of a single gene but of a *set of genes*. To denote a genotype let us separate by a semicolon the symbols which stand for its maternal and paternal heritage. Thus a genotype may be denoted by $(x; y)$ where the letter before (after) the semicolon denotes the individual's maternal (paternal) heritage. Also introduce $w^{(n)}(x; y)$, the distribution of genotypes, and assume that, as in (1) and (2):

$$(x; y) = (y; x) \tag{1'}$$

and

$$w^{(n)}(x; y) = w^{(n)}(y; x). \tag{2'}$$

This differentiation with respect to the maternal and paternal heritage is of great importance for the understanding of linkage. It relates however merely *to the formation of new gametes.*

In the example under consideration where we are concerned with the study of m loci, "x" stands for the "maternal" genes, $x_1, \cdots x_m$, while "y" symbolizes the m "paternal" genes, $y_1, \cdots y_m$. The kinds of gametes which this organism may produce depend on the possible combinations of the material it has inherited. These combinations happen according to a probability law which we again call the segregation distribution. It is one of the main tasks of the theory to define a segregation distribution which corresponds to given biological conditions. In a recent paper [5, d], R. A. Fisher says: "The laws of inheritance are the rules whereby, given the constitution of an organism, the kinds of gametes it can produce and their relative frequencies can be predicted." Next, the distribution of gametes $p^{(n)}(z)$ is derived from the distribution of genotypes by means of the segregation distribution. Finally, the random fusion of the two gametes, expressed in the formula $w^{(n+1)}(x; y) = p^{(n)}(x)p^{(n)}(y)$, completes the cycle of inheritance.

It is easy to indicate a few mathematical problems which present themselves in this and in similar situations:

a) The possible genotypes are to be completely and simply enumerated under consideration of the specific biological conditions which characterize a given situation. The same problem exists for the various kinds of gametes.

b) An adequate segregation distribution is to be defined.

In a), as well as in b), the mathematical approach can help to simplify and clarify the concepts. As in other sciences the actual situation in nature is often so complex that "models" have to be constructed representing a compromise between mathematical simplicity and biologi-

cal adequacy. By such a procedure we introduce for example in mechanics models like the "rigid body," the "elastic body," the "ideal fluid," etc.

c) The distribution of gametes, $p^{(n)}(z)$ is to be derived from $w^{(n)}(x; y)$ by means of the segregation distribution, and *recurrence relations* for the $p^{(n)}(z)$ are desired. They are in general, simpler than those for the $w^{(n)}(x; y)$; under random mating both are equivalent.

d) We wish to *integrate* these recurrence equations which are in general, non linear difference equations with constant coefficients; that means we want to express $p^{(n)}(z)$ in terms of "n," of the initial values $p^{(0)}(z)$,—which in turn have to be derived from the $w^{(0)}(x; y)$,—and of the parameters involved in the definition of the segregation distribution.

e) We want to know whether one or more *states of equilibrium* prevail for $p^{(n)}(z)$ and for $w^{(n)}(x; y)$ and if so under what conditions. To settle this we also have to investigate the *limit behaviour* of $p^{(n)}$ and $w^{(n)}$ as "n" tends towards infinity.

In the following we shall comment on the problems a)–e) in the case of: 1) *m loci* (*linkage*), assuming autosomal genes and "diploids"; 2) so-called *autopolyploids* (as opposed to "diploids") for one as well as for several loci.

4. LINKAGE IN THE CASE OF m LOCI AND r ALLELES

As in the first example of the preceding section assume m loci. Mendel's original assumption was that all possible gametic combinations are equally probable. This assumption of *independent assortment* does not introduce any parameters since each of the 2^m possible combinations appears with probability $1/2^m$.[1]

This simple conception was shaken by the observation of "linkage" associated with the names of Bateson and Punnet, of Morgan, Sturtevant, and of many other well known biologists of our time. It appeared that not all possible gametes occur with equal frequency. Consider for a moment the case $m=2$, and r alleles. We denote a type by $(x_1x_2; y_1y_2)$ where x_i as well as y_i, $(i=1, 2)$, relate to the ith locus and each of them takes on the values $1, 2, \cdots r$. The four possible types of gametes are then x_1x_2, x_1y_2, y_1x_2, y_1y_2, since one gene is transmitted with respect to each locus. In Mendel's conception each of these four possibilities has the probability $\frac{1}{4}$. In modern genetics these four probabilities are

[1] It is referred to by many authors as "random segregation." This term should not be confused with the general use of the term "random" underlying each probability distribution and not merely the particular one where all probabilities are equal. Such a probability distribution is often called "uniform," in probability calculus.

assumed to be $(1-c)/2$, $c/2$, $c/2$, $(1-c)/2$, where c is the so-called "recombination value." It is often assumed by biologists that c is less than $\frac{1}{2}$. This implies that the genes that came in together exhibit a tendency to stay together; mathematically, however, c may have any value between zero and one. The introduction of the parameter c corresponds to assuming a certain influence of the grand-parents. The case, $m=r=2$, has been completely investigated by Weinberg [25], Robbins [20] and Jennings [13, b, c].

In the general case there are $m(m-1)/2$ recombination values, c_{ij}. These may be defined in a way which is independent of additional assumptions like "chiasma theory," or "linear theory": c_{ij} ($i=1, \cdots m$, $j=1, \cdots m$, $i\neq j$) is the probability that those transmitted genes whose subscripts are i and j come from different parents, or, in other words, the probability that either x_i and y_j or x_j and y_i be transmitted, no matter what happens to the $(m-2)$ other factors. It follows that c_{ij} is a marginal probability of our segregation distribution. It will be seen that, unless we admit some additional biological hypothesis, the $m(m-1)/2$ recombination values are not sufficient for the description of general linkage.

In case of m loci a genotype is characterized by *two m-dimensional vectors*, x_i and y_i ($i=1, \cdots m$, $x_i=1, \cdots r$, $y_i=1, \cdots r$). This simple idea, absolutely basic in the author's approach, is not generally accepted. Often (the consideration being limited to two or three alleles) a type is simply denoted by m pairs of numbers, which is all right only if independent assortment is assumed. Then there are just nine types for $m=r=2$ and in general $[r(r+1)/2]^m$ types. For the understanding of linkage, however, the assumption of two sets, each of m numbers, is necessary. This amounts to distinguishing for example between $(AB; ab)$ and $(Ab; aB)$; in fact, these two types *are* different with respect to the segregation of gametes if $c\neq\frac{1}{2}$, since the first transmits AB with probability $(1-c)/2$, and the second transmits AB with probability $c/2$. There are in this case ten and not nine types and in the general case there are $\frac{1}{2}r^m(r^m+1)$ types.

Thus, a type is denoted by $(x_1, \cdots x_m; y_1, \cdots y_m)$. In the formation of a gamete a new set of m elements is composed in such a way that corresponding to each of the m subscripts either the x-value or the y-value is chosen. Consequently there are 2^m possibilities. The gamete will consist of some of the x_i and of some of the y_i. By fusion of this gamete with another, a zygote is formed which contains in a new combination some of the material found in the parents. This new combination, again, will not persist: when the new being forms sex cells the

new gamete will exhibit a new characteristic combination and the old combinations disappear. Thus the genes recur from generation to generation passing through many individuals which they determine (with respect to certain properties). The combinations change, but not the constituent genes. (This is a very simplified scheme which does not even cover mutations.)

To describe adequately the 2^m possible assortments we introduce a segregation distribution which, in this case, may be called a *linkage distribution.* Denote by S the set of numbers, 1, 2, $\cdots m$, by T any subset of S, $(O \leqq T \leqq S)$ and by l_T *the probability that the maternal genes belonging to T and the paternal genes belonging to $T' = S - T$ be transmitted.* Note that this definition is independent of the genotype of the parent. There are 2^m such probabilities. For obvious symmetry reasons we must assume

$$l_T = l_{T'}. \tag{7}$$

Hence, since the sum of all probabilities is *one,* we have introduced

$$M = 2^{m-1} - 1$$

parameters. For $m \geqq 4$, $M > m(m-1)/2$; so there are in general more linkage-parameters than recombination values. It is desirable to describe linkage in terms of the $m(m-1)/2$ recombination values, that is, to express the M linkage parameters by these values. This problem is closely connected with the so-called "linear arrangement" of the genes of the same linkage group and with the concept of "distance" of linked genes. R. A. Fisher [5, e] recently offered a very suggestive solution. An older solution, fairly generally accepted (but in the author's opinion open to criticism) is due to Haldane [9, c]. Whatever the relation between linkage distribution and recombination values, the linkage distribution must be assumed known if we wish to study the heredity problem of linked genes.

5. SOLUTION OF THE LINKAGE PROBLEM

For the above linkage problem, the author has solved the problems indicated in section 3. We have considered a) the enumeration of the possible genotypes under linkage and b) the definition of the linkage distribution. Before continuing, some important particular forms of the linkage distribution should be specified: 1) In the case of *independent assortment* (random segregation), the 2^m values of the linkage distribution are all equal to each other, hence each equal to $1/2^m$. 2) If there are

S distinct "*linkage groups,*" the m-dimensional linkage distribution[2] resolves into the product of S probability distributions. 3) The m loci may form several groups of *completely linked loci.* Then, all recombination values within each group are zero while all recombination values between members of any two different groups equal $\frac{1}{2}$.

The main problem is the *recurrence problem* (problem c) of sec. 3) which always occupies the central place. In the case of m arbitrarily linked (autosomal) genes it has been solved [7a], [13b]. To explain the solution we need the concept of a "marginal distribution," well known to statisticians. If $p(z_1, z_2, \cdots z_m)$ denotes a discrete (arithmetical) probability distribution in m variates, then $p_{12}(z_1z_2) = \sum_{z_3} \cdots \sum_{z_m} p(z_1z_2z_3 \cdots z_m)$ is the probability of the result (z_1z_2) and similarly $p_{ij}(z_iz_j)$, $p_i(z_i)$ or $p_{ijk}(z_iz_jz_k)$ may be defined. Note that there are marginal distributions of "order 1," of "order 2," etc. Obviously we may write $p_T(z_T)$ for $p_{ijk}(z_iz_jz_k)$ if T denotes the set (i, j, k) and z_T the set $(z_iz_jz_k)$. If, finally, we write $p^{(n)}(z)$ for $p^{(n)}(z_1z_2 \cdots z_m)$ our recurrence formula reads

$$p^{(n+1)}(z) = \sum_{(T)} l_T p_{T'}(z_T) p_T;(z_{T'}) \tag{9}$$

where the sum is over all subsets T of S and $p_T(z_T)$ is the marginal distribution whose subscripts are the points of T. We have e.g. (anticipating the result (11):

$$m = 2: \quad p^{(n+1)}(z_1z_2) = 2l(00)p^{(n)}(z_1z_2) + 2l(01)p^{(0)}(z_1)p^{(0)}(z_2)$$

$$\begin{aligned} m = 4: \; p^{(n+1)}(z_1z_2z_3z_4) = 2\{&l(0000)p^{(n)}(z_1z_2z_3z_4) \\ &+ [l(1000)p^{(0)}(z_1)p^{(n)}(z_2z_3z_4) \\ &+ \cdots [l(0001)p^{(0)}(z_4)p^{(n)}(z_1z_2z_3)] \\ &+ [l(1100)p^{(n)}(z_1z_2)p^{(n)}(z_3z_4) \\ &+ l(1010)p^{(n)}(z_1z_3)p^{(n)}(z_2z_4) + \cdot \;]\}. \end{aligned}$$

In our recurrence formulas the values of the linkage distribution act as essential "separators" between meaningful groups of probabilities. This clear and simple recurrence relation is characteristic for "random mating" and "chromosome segregation" (see also next section).

The next problem consists in the *solution of the recurrence equations.*

---

[2] A definition of the linkage distribution, completely equivalent to the previous one, is the following: There are 2^m probabilities with sum *one*, $l(\epsilon_1, \epsilon_2, \cdots \epsilon_m)$ where ϵ_i equals either zero or one. $\epsilon_i = 1$ means that the transmitted gene corresponding to the ith factor is the maternal one while $\epsilon_i = 0$ means that the paternal value has been transmitted. Instead of (7) we have

$$l(\epsilon_1\epsilon_2 \cdots \epsilon_m) = l(1-\epsilon_1, 1-\epsilon_2, \cdots, 1-\epsilon_m)$$

These form a system of *quadratic difference equations with constant coefficients*. It has been shown that this solution can be found [7a]; it is however no simple matter so long as everything remains unspecified. In the case of independent assortment a very elegant solution has been given by H. Tietze [24] who established the recurrence relations for this case and investigated the limit behaviour in full generality. Jennings [13, b, c] and Robbins [20] have solved the general linkage problem for $m=2$, where one parameter enters, including an explicit expression for $p^{(n)}$. In the case $m=3$, where three parameters are involved, the explicit solution given by the author [7b] is still very simple.

Unless we have *complete linkage* ($l(1, \cdots, 1)=l(0, \cdots, 0)=\frac{1}{2}$, all other linkage probabilities zero) there is *no equilibrium for finite n*. The investigation of the *limit behaviour* of $p^{(n)}(z)$, as $n\to\infty$ becomes important. The author has established [7a] general results which may be formulated in the following statements:

First it is shown that, just as in the case $m=1$, *the gametic proportion of each single gene remains constant through the generations*

$$p^{(n)}(z_i) = p^{(0)}(z_i) \qquad (i = 1, 2, \cdots, m, n = 0, 1, \cdots). \tag{11}$$

This is the mathematical expression of the "immortality" of the genes. Next, consider the joint distribution $p^{(n)}(z)$. There are two important particular cases which have to be settled first. *Assume* $l(00 \cdots 0)=\frac{1}{2}$; *then all recombination values equal zero*, or, in other words, all "mixed" gametes, containing partly paternal and partly maternal genes, are a priori excluded. *We then have for all n:*

$$p^{(n)}(z_1 \cdots z_m) = p^{(0)}(z_1 \cdots z_m). \tag{12}$$

Completely linked genes act like one. The other extreme is *that no recombination value equals zero. Then it follows that*

$$\lim_{n\to\infty} p^{(n)}(z_1 \cdots z_m) = p^{(0)}(z_1) \cdots p^{(0)}(z_m). \tag{13}$$

In Mendel-Tietze's [24] case of independent assortment each $c_{ij}=\frac{1}{2}$, hence the above condition is satisfied and (13) holds.

Finally, if $c_{ij}>0$ does not hold for *all* pairs i, j, the linkage distribution degenerates in various ways into $t\leq m$ *groups of complete linkage.* By that we mean a set of genes $(i, j, \cdots, k)$ *within which no recombination takes place*, i.e. all corresponding c_{ij} are zero. For example if (1, 3, 6) is such a set, it can be shown that the *marginal distribution* $p_{136}{}^{(n)}$ *is preserved through the generations*: $p_{136}{}^{(n)}=p_{136}{}^{(0)}$ *holds for all n*, besides (11). To describe the limit behaviour completely we use the term:

maximal set of completely linked genes for a subset T of S containing *as many as possible of the m numbers with no recombination within the set.* Assume e.g. $m=8$, and that the group (1, 4) as well as the group (2, 3, 5, 6) are each completely linked, while the recombination values c_{12}, c_{17}, c_{18}, c_{27}, and c_{28} are known to be different from zero. Then (14), (2356), (7), (8) are the maximal sets of completely linked genes and, in an obvious notation:

$$\lim_{n\to\infty} p_{12\cdots 8}{}^{(n)} = p_{14}{}^{(0)} p_{2356}{}^{(0)} p_7{}^{(0)} p_8{}^{(0)} \tag{14}$$

or, more generally [7, b]:

If $S_1, \cdots, S_t$ are the maximal sets of completely linked characters ($t \leqq m$), each containing at least one element, then

$$\lim p_{12\cdots m}{}^{(n)} = p_{S_1}{}^{(0)} \cdots p_{S_t}{}^{(0)}. \tag{15}$$

The general theorem may be expressed in a more complete way:

Under the conditions of our investigation (m arbitrarily linked genes, random mating, etc.): *For the gene distribution in successive generations the original* (marginal) *distributions of each "maximal group" S_t are preserved: $p_{S_t}{}^{(n)} = p_{S_t}{}^{(0)}$, for all n; as $n \to \infty$, the joint distribution $p^{(n)}(z)$ of the m random variables approaches a limit where the different sets of genes are independently distributed.*

It is obvious that from these results corresponding results for the $w^{(n)}(x; y)$ follow easily, by means of the equation which is the analogue of (6).

It is biologically important and mathematically interesting to consider the extension where *different linkage distributions for males and females* are assumed. At the suggestion of S. Wright, who considered the problem for $m=2$ and $m=3$, the author has investigated this extension [7, d], assuming two linkage distributions l and l'. It turns out that the recurrence relations change, but not in an essential way: the arithmetic means of the values of the two linkage distributions play a decisive role. There appear now two different distributions of gametes corresponding to the two sexes, and simple recurrence relations hold for their arithmetical mean. As $n \to \infty$ the *two* gametic distributions approach *the same limit* which is independent of the linkage distributions and where the m genes are independently distributed.

6. THEORETICAL GENETICS OF AUTOPOLYPLOIDS

Thus far we have discussed the extension of Mendel's original ideas

to the case of linkage. A second remarkable extension has come with the discovery of *polysomic inheritance*, as opposed to normal or diploid inheritance. We shall discuss the important and interesting case of *autopolyploids*. This problem has been studied by numerous biologists. Among them we mention, because of the more theoretical aspect of their work, R. P. Gregory, H. J. Muller, J. B. S. Haldane [2], [9 e], S. Wright [26, c], K. Mather [16, b, c], [6], and R. A. Fisher [5, b, c, d].

In the preceding pages we assumed "chromosome segregation" as opposed to "chromatid segregation." It can be shown easily that for normal or diploid organisms, as considered so far, there is, mathematically, no difference between chromosome segregation and chromatid segregation. Thus the results of section 5 are general. This situation changes for polyploids, where there is a definite difference between the two modes of segregation. Chromosome segregation may be considered as an approximation to chromatid segregation; this latter one actually prevails according to modern studies. Nevertheless we shall discuss mainly chromosome segregation, since even under this simpler assumption the problem of polyploids appears rather complicated and unfamiliar to the statistician.

In the simplest case where only *one* locus is considered, an autopolyploid organism may be described as follows: In case of a $2s$ ploid each gamete consists not of *one* but of $s \geqq 1$ genes. A genotype possesses *two sets, each of s genes*, each gene being represented by one of the r numbers $1, 2, \cdots, r$, the r alleles, where r may be greater than, less than, or equal to s. Although a genotype is mathematically described by two sets, each of s numbers, there are less than $N = r^{2s}$ distinct types since many types have to be considered as equal. First, we have as before: $(x; y) = (y; x)$. Moreover, it is assumed with respect to the maternal (paternal) heritage that there is no difference between the various permutations of the s numbers which constitute this heritage. In accordance with this, a type is denoted, e.g. for $s = 5$, by $(a_1^2 a_2^3 : a_1 a_4 a_6 a_7^2)$. It is then easily seen that there are

$$(16) \qquad N_1 = \tfrac{1}{2}R(R+1) \text{ genotypes where } R = \binom{r+s-1}{s}.$$

With respect to segregation the results of observations suggest the following: In the formation of a new individual each parent transmits to the offspring a set of s genes out of the $2s$ genes the parent possesses. The selection of these s genes happens according to a probability dis-

tribution called the segregation distribution, which need not be the same for males and females. We assume here, however, that there is one and the same segregation distribution for both sexes. Moreover, the segregation distribution is independent of the genotype of the parent and of n. Let us consider its definition. Out of $2s$ numbers a set of s numbers can be selected in S ways, where

$$S = \binom{2s}{s} = \frac{(2s)!}{s!s!}$$

These S cases (there are six in case of the most often considered tetraploid, $s=2$) have been more or less tacitly assumed equally probable, first by Muller, then, as far as I know, by all biologists concerned with the problem. This assumption is not logically necessary and does not seem inevitable in the light of reported observations. We therefore introduce a segregation distribution where the S probabilities are not a priori assumed equal. On the other hand we must avoid introducing parameters that are not meaningful biologically. After considering the analogy to other linkage phenomena and studying the numerical results of observations, the author [7c] was led to the hypothesis that within the s transmitted genes the *proportion* of maternal and paternal genes plays a certain role. We thus make the following definition: Call

$$(17) \qquad l_\alpha = \binom{s}{\alpha}\lambda_\alpha \qquad (\alpha = 0, 1, \cdots, s)$$

the probability that a specified set of α maternal genes be transmitted. Assume symmetry of paternal and maternal heritage:

$$(18) \qquad l_\alpha = l_{s-\alpha}.$$

Since a set of α specified maternal genes may be combined with $(s-\alpha)$ paternal genes in $\binom{s}{\alpha}$ ways, we have

$$(19) \qquad \sum_\alpha^{0\cdots s} \binom{s}{\alpha} l_\alpha = \sum_\alpha^{0\cdots s} \binom{s}{\alpha}^2 \lambda_\alpha = 1.$$

Thus if $s=2\mu$ or $2\mu+1$, just μ parameters are introduced. These μ parameters may be partly or all equal to each other. If $\lambda_\alpha = 1/S$ we have "random chromosome segregation" (Muller, Haldane). Other particular assumptions may be considered.

Although there is a certain apparent similarity between this mathematical formulation and that in the linkage problem, the situations are in fact quite different. In the polyploid problem, any α maternal

genes may be segregated together with any $(s-\alpha)$ paternal ones, while in the linkage problem the new set of m genes has to contain exactly one value, either the x_i or the y_i, with respect to each of the m subscripts, $1, 2, \cdots, m$.

Our three basic distributions, the distributions of genotypes and of gametes and the segregation distribution have thus been defined. The next step, and the most important one, consists in finding a general recurrence formula which permits us to derive step by step the $p^{(n)}$ for $n=1, 2, \cdots$, starting with $p^{(0)}$. It turns out that there exists for polyploids too a surprisingly simple recurrence relation. It is based on the above concepts and on the use of certain types of "marginal distributions" whose definition is much less obvious than in the linkage problem; this recurrence law holds for any "r" and any "s." For a homozygotic gamete of type (A^s) our recurrence formula is simply

$$p^{(n)}(A^s) = \sum_{\alpha}^{0\cdots s} l_\alpha \cdot \binom{s}{\alpha} p^{(n)}(A^\alpha) p^{(n)}(A^{s-\alpha}). \tag{20}$$

The $p^{(n)}(A^\alpha)$ and $p^{(n)}(A^{s-\alpha})$ are marginal distributions defined in agreement with the general concept of a marginal distribution. For example $p^{(n)}(A^\alpha)$ is the probability of a gamete with α A-genes, no matter what its remaining $(s-\alpha)$ genes may be. If we consider heterozygotes, r different allelomorphs, where $r \gtreqless s$, so that a gamete is of type $(a_1^{x_1} \cdots a_r^{x_r})$ with $x_1 + \cdots + x_r = s$, we obtain a result which is not much more complicated than (20).

By means of these recurrence relations the author [7c] has derived a *limit theorem*, as $n \rightarrow \infty$. Haldane [9e] has indicated a distribution which reproduces itself under random segregation and thus represents a state of equilibrium. This result leaves open the biological and mathematical question of whether such an equilibrium is actually reached, and if so, under what conditions. Our limit theorem may be formulated as follows:

Denote the r alleles by $a_1, \cdots, a_r$, and consider one locus, and chromosome segregation. Then

1) *As in case of a diploid* (sec. 1) *the gametic proportions corresponding to each single allele remain unchanged through the generations,* $p^{(n)}(a_i) = p^{(0)}(a_i)$, $(i=1, \cdots, r, n=0, 1, \cdots,)$.

2) *If and only if $l_0 < \frac{1}{2}$ (i.e. if "mixed gametes" are not a priori excluded) the joint distribution of gametes converges towards a limit where the alleles are independently distributed.*

$$\lim_{n\rightarrow} p^{(n)}(a_1^{x_1} \cdots a_r^{x_r}) = [p^{(0)}(a_1)]^{x_1} \cdots [p^{(0)}(a_r)]^{x_r}. \tag{21}$$

In the case of Haldane's "random segregation" (where all segregation probabilities are equal to each other), the condition of our limit theorem is satisfied and our result thus proves and completes his statement.

While chromatid segregation describes the biological situation more correctly than chromosome segregation it seems that this latter, with an appropriate general segregation distribution, presents a useful approximation. It would lead us too far to attempt an explanation of chromatid segregation. A few remarks must suffice. Chromosome segregation is a particular case of chromatid segregation, where certain probabilities, corresponding to "double reduction," are assumed to equal zero. A segregation distribution for chromatid segregation has been introduced by K. Mather and R. A. Fisher [6]. This basic approach has been generalized by the author [7, f] so as to contain as particular cases Fisher-Mather's segregation distribution [6], Haldane's "random chromatid segregation" [9, e], and my general chromosome segregation distribution [7, c]. These investigations are for $s=2$, 3, and 4. They include (according to the program outlined in section 3) *recurrence relations*; *solution* of those difference equations; and *limit-results*. It is worth mentioning that in this case the recurrence relations, as well as the limit theorems, are *essentially different*—and this in a very interesting way—from the results (20) and (21). While these last are of the same type as the corresponding results (9) and (13), the problems of chromatid segregation introduce an entirely new type of statistico-biological laws.

The next step leads to the study of *polyploids under consideration of linkage* (several loci). In a particular case (data from tetraploid primulas), DeWinton and Haldane [2] have proposed a linkage theory for tetraploids under chromosome segregation. A suggestively simple segregation distribution is introduced on the basis of evidence which indicated a certain "pairing" as prevailing in this plant. In a more recent paper Fisher [5, d] considers linkage under chromatid segregation. His paper is concerned with the enumeration of gametes and of genotypes and to a certain extent with the definition of a segregation distribution, problems which in this general form are far from easy.

The formulation of the problem of $2s$-ploids with m loci must include the problem of diploids for m loci (sections 4 and 5) as well as that of a $2s$-ploid for one locus (section 6). The general problem must reduce to these problems and respective results if either $s=1$, or $m=1$. Further, the case of a $2s_1$-ploid with m_1 loci ($s_1 \leqq s$, $m_1 \leqq m$) must appear as a "marginal" case of the more general one. Accordingly, I introduce [7, e]

a sufficiently general segregation distribution. Recurrence formulas are also derived and the limit behaviour of the characteristic distributions is completely investigated. The results are simple and complete. It seems that these structurally clear recurrence formulas describe relevant features of a rather general biological situation: $2s$-ploid, m loci, r alleles, random mating, chromosome segregation,—with Mendel's theory of heredity as the basis. A similar remark holds for the limit theorems [7f].

The study of *linkage of polyploids under chromatid segregation* seems to be very difficult. The aim is, of course, to study the problem by means of a segregation distribution which is well adapted to the available data and which satisfies the theoretical requirements. With such a distribution at the basis we want to obtain a fairly complete insight into the theoretical side of the problem, including recurrence relations, stability, limit behaviour, and rate of approach to equilibrium. The conclusions reached so far for $m=1$ leave no doubt that we have to expect an entirely new type of results.

7. SOME OPEN PROBLEMS

The mathematical description of polyploid linkage under chromatid segregation constitutes an example of an open problem. There are many open problems within the limits set by the title of this paper. Let us mention a few examples.

The genetics of autopolyploids should be investigated under the assumption of different segregation distributions for the two sexes. In the vast domain of selection problems the results are still incomplete in many instances, even from a purely mathematical point of view. As a very simple example consider the first of the selection problems described in section 2. Here, a general integration has not been given in case of arbitrary selection coefficients; the limit behaviour of the distributions, however, can be completely described (cf. in particular Haldane's [9d] basic work on selection). If we assume different selection coefficients for the two sexes, then, to the author's knowledge, not even the limit behaviour has been investigated for the general case. The generalization from two to r alleles, almost trivial in case of random mating, introduces considerable difficulties in selection problems. Linkage under differential-viability-selection as well as polysomic inheritance are not easy problems. There are similar open problems for the various "systems of mating" which involve some choice of the mate. These examples have been chosen such as not to require the explanation of new biological situations. Other mathematical problems arise in

connection with so-called "sex-linked" inheritance in contrast to "autosomal" inheritance which has been considered throughout this paper. If the assumption of "non-overlapping generations" (see section 1) is dropped we are faced with completely new and quite difficult problems.

The results reported here and the problems explained may appear, even to the mathematically-minded biologist, as rather formal and somehow remote from his field of interest. Modern genetics is an extremely complex science, with relations to various fields of knowledge, such as general biology, biophysics and biochemistry, general anthropology, physiology, and psychology. "Formal genetics" as considered in this paper represents only one side of a manysided problem. We need not however lose sight of the totality of a problem if we follow up thoroughly only one aspect. Modern physics began when Galileo performed in a scientific way some very simple experiments and described them in mathematical terms; modern genetics started with Mendel's exact observations of simple phenomena and their mathematical description. Today, in biology as well as in other branches of knowledge, the mathematical approach hardly needs justification. Nevertheless the biologist may sometimes wonder whether the generalizations which the mathematician uses are of any value to him. Each kind of approach has its inherent logic and must abide by its own inner laws. The mathematical approach to biological problems, by way of abstraction and generalization, will prove rewarding at present and in the future.

SELECTED REFERENCES

[1] F. Bernstein
  (a) "Variations- und Erblichkeitsstatistik," *Handbuch der Vererbungswissenschaft*, Bd. 1, pp. 1–96.
  (b) "Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen," *Zs. f. Induktive Abstammungs u. vererbungslehre*, Vol. 37, (1925).

[2] D. De Winton and J. B. S. Haldane, "Linkage in the tetraploid Primula Sinensis," *Journal of Genetics*, Vol. 24 (1931), pp. 1–44.

[3] Th. Dobzhansky, *Genetics and the Origin of Species*, Columbia Univ. Press, (1938).

[4] Th. Dobzhansky and L. C. Dunn, *Heredity, Race, and Society*, Penguin Books, New York, 1946.

[5] R. A. Fisher
  (a) *The Genetical Theory of Natural Selection*, Oxford Press, 1930.
  (b) "The theoretical consequences of polyploid inheritance for the third style form of Lythrum salicaria," *Annals of Eugenics*, Vol. 11 (1941), pp. 31–39.
  (c) "Allowance for double reduction in the calculation of genotype frequencies with polysomic inheritance," *Annals of Eugenics*, Vol. 12 (1944), pp. 169–171.

(d) "The theory of linkage in polysomic inheritance," *Phil. Trans. Roy. Soc. London*, B, 233 (1947), pp. 55–87.
(e) "A quantitative theory of genetic recombination and chiasma formation," *Biometrics*, Vol. 4 (1948), pp. 1–13.

[6] R. A. Fisher and K. Mather, "The inheritance of style length in Lythrum Salicaria," *Annals of Eugenics*, Vol. 12 (1943), pp. 1–23.

[7] H. Geiringer
(a) "On the probability theory of linkage in Mendelian heredity," *Annals of Mathematical Statistics*, Vol. 15 (1944), pp. 25–57.
(b) "Further remarks on linkage in Mendelian heredity," *Ibidem*, Vol. 16 (1945), pp. 390–398.
(c) "Contribution to the heredity theory of multivalents," *Journal of Mathematics and Physics*, Vol. 26 (1948), pp. 246–278.
(d) "On the mathematics of random mating in case of different recombination values for males and females," *Genetics*, 33 (1948), pp. 548–564.
(e) "Contribution to the linkage theory of autopolyploids," I and II, *Bulletin Mathem. Biophysics* Vol. 11, (1949), I, pp. 59–82, and II, pp. 197–219.
(f) "Chromatid segregation of tetraploids and hexaploids," forthcoming *Genetics*, 34 (1949).

[8] H. Geppert and S. Koller, *Erbmathematik.* Leipzig 1938.

[9] J. B. S. Haldane
(a) *The Causes of Evolution*, London, New York, 1932.
(b) *New Paths in Genetics*, London, 1941.
(c) "The combination of linkage values and the calculation of distances between the loci of linked factors," *Journal of Genetics*, Vol. 8 (1919), pp. 299–308.
(d) "A mathematical theory of natural and artificial selection, *Transactions of the Cambridge Philosophical Society*, Vol. 23 (1924), pp. 19–41. Successive parts in later volumes: *Proceedings, Cambridge Philosophical Society*, 23 (1927), pp. 363–372, 607–615, 838–848, *ibidem*, 26 (1930), pp. 220–230, 27 (1931), pp. 131–142, 28 (1932), pp. 244–248.
(e) "Theoretical genetics of autopolyploids," *Journal of Genetics*, Vol. 22 (1930), pp. 359–372.
(f) "The rate of spontaneous mutation of a human gene," *Journal of Genetics*, Vol. 31 (1935).
(g) "Some problems of mathematical biology," *Jour. of Mathematics and Physics*, Vol. 14 (1935), pp. 125–136.

[10] G. H. Hardy, "Mendelian proportions in a mixed population," *Science*, Vol. 28 (1908), pp. 49–50.

[11] A. Heilbronn and K. Kosswig, "Principia Genetica," *The Journal of Unified Science*, Vol. 8 (1939), pp. 229–255.

[12] L. Hogben, *An Introduction to Mathematical Genetics*, W. W. Norton, New York, 1946.

[13] H. S. Jennings
(a) "The numerical results of diverse systems of breeding," *Genetics*, Vol. 1 (1916), pp. 53–89.
(b) "The numerical results of diverse systems of breeding with respect to two pairs of characters etc. etc.," *Genetics*, Vol. 2 (1917), pp. 97–154.

(c) "The numerical relations in the crossing over of the genes etc., etc.," *Genetics*, Vol. 8 (1923), pp. 393–457.

[14] W. Johannsen, *Elemente der exacten Erblichkeitslehre*. 2. Auflage, Jena, 1913.

[15] V. A. Kostitzin, *Biologie Mathématique*. (Préface de Vito Volterra) Paris, 1937.

[16] K. Mather

(a) *The Measurement of Linkage in Heredity*. London, 1938.

(b) "Reductional and equational separation of the chromosomes in bivalents and multivalents," *Journal of Genetics*, Vol. 30 (1935), pp. 53–77.

(c) "Segregation and linkage in autotetraploids," *Ibidem*, Vol. 32 (1936), pp. 287–314.

[17] G. Mendel, "Versuche über Pflanzenhybriden," *Verhandlg. d. Naturforschenden Ver. in Brünn*. Abhandlungen, IV. Bd., Brünn, 1866, pp. 3–47.

[18] T. H. Morgan, *The Theory of the Gene*, New Haven, 1928.

[19] K. Pearson, "On a generalized theory of alternative inheritance with special reference to Mendel's law," *Phil. Trans. Roy. Soc.* (A), Vol. 203 (1904), pp. 53–86.

[20] R. B. Robbins, "Applications of mathematics to breeding problems II, and III," *Genetics*, Vol. 3 (1918), pp. 73–92 and pp. 375–389.

[21] F. W. Sansome, "Chromatid segregation in Solanum Lycopersicum, *Journal of Genetics*, Vol. 27 (1933), pp. 105–132.

[22] F. W. Sansome and J. Philp, *Recent Advances in Plant Genetics*, Sec. Ed., London, 1939.

[23] L. H. Snyder and C. W. Cotterman, "Studies on Human Inheritance XVII, Gene frequency analysis etc.," *Genetica*, Vol. 19 (1937), pp. 537–552

[24] H. Tietze, "Über das Schicksal gemischter Populationen nach den Mendelschen Vererbungsgesetzen," *Zs. Angew. Math. u. Mech.*, Bd. 3 (1923), pp. 362–393.

[25] W. Weinberg, "Über Vererbungsgesetze beim Menschen," *Zs. f. Indukt. Abstammungs- und Vererbungslehre*, Vol. 1 (1909), pp. 277–330.

[26] S. Wright

(a) "Systems of mating. I–V," *Genetics*, Vol. 6 (1921), pp. 111–178.

(b) "Evolution in Mendelian populations," *Genetics*, Vol. 16 (1931) pp. 97–159.

(c) "The distribution of gene frequencies in populations of autopolyploids," *Proc. Nat. Acad. Sciences*, Vol. 24 (1938), pp. 372–377.

(d) "Statistical genetics and evolution," *Bulletin of the American Math. Soc.*, Vol. 48 (1942), pp. 223–246.

(e) "Isolation by distance under diverse systems of mating," *Genetics*, Vol. 31 (1946), pp. 39–59.

(f) "On the roles of directed and random changes in gene frequency in the genetics of populations." *Evolution*, Vol. II, (1948), pp. 279–294.

THE FITTING OF LOGISTIC CURVES BY MEANS OF A NOMOGRAPH

EUGENE A. RASOR
Federal Security Agency, Social Security Administration

"Growth" curves, such as logistic curves, are widely used. This article discusses a previous paper on this subject and shows how simpler nomographs for fitting the logistic curve may be developed. First, a simpler nomograph of the same type as in the previous article is presented, and then a different type is developed which has the advantage of practically no computations, simplicity of appearance, and a wider range of values.

SPURR AND ARNOLD have shown how to fit a logistic curve quite simply by means of a nomograph.[1] This paper will first present a nomograph for finding the upper asymptote determined in the same fashion as did Spurr and Arnold, but somewhat simpler in having two of the scales parallel, and in addition taking in more values. Next, this paper will develop a nomograph of a different type, by means of which the logistic curve may be more readily and more accurately fitted to three equidistant values.

Chart 1 shows the nomograph for determining the upper limit of the logistic curve,

$$y = k/(1 + e^{a+bx}),$$

with the three scales being y_1/y_0, y_2/y_1, and k/y_0. It will be noted that the latter two scales are parallel and cover roughly the same values as did the Spurr-Arnold nomograph. However, the other scale shows a range of from 1.4 to 30.0, as contrasted with only 1.5 to 4.2 in the Spurr-Arnold nomograph. Thus, not only is the nomograph in Chart 1 more symmetrical, but also by utilizing the entire sheet it can cover more possible cases.

---

[1] William A. Spurr and David R. Arnold, "A Short-Cut Method of Fitting a Logistic Curve," *Journal of the American Statistical Association*, March 1948.

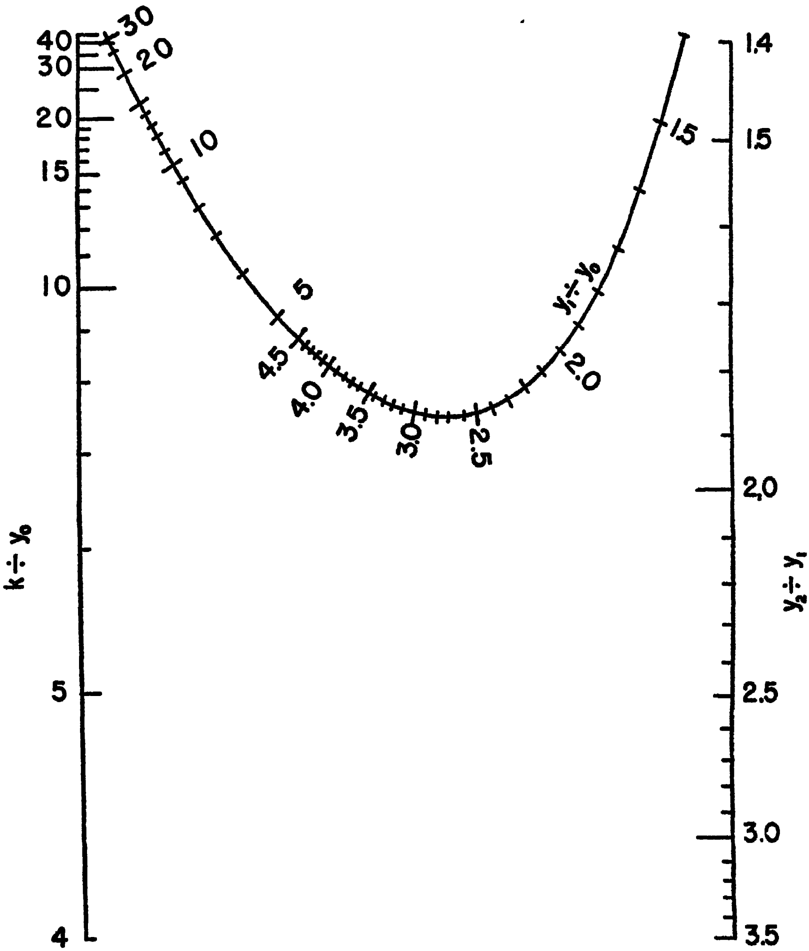


CHART 1. NOMOGRAPH FOR DETERMINING k IN $y = \frac{k}{1+e^{a+bx}}$

This nomograph has been determined from the following determinant:

$$\begin{vmatrix} 0 & \dfrac{44(\theta-4)}{41(\theta-1)}\mu & \\ & \dfrac{(7-2B)}{3B} & \\ \dfrac{396A\delta}{287A^2-178A+287} & \dfrac{44(7A^2-20A+28)}{287A^2-178A+287}, & 1 \end{vmatrix} = 0$$

where $A=y_1/y_0$, $B=y_2/y_1$ and $\theta=k/y_0$ as in the article by Spurr and Arnold; δ and μ are the width and height respectively of the chart.

Next, turning to the development of a simpler nomograph which will give accurate values of k, a, and b, directly from three equidistant values of y, without any preliminary calculations of the ratios of the y's and subsequent multiplication, along with intermediate (or extended) values of y_x directly from the graph, we have the following development:

For example, given the three equidistant values, y_0, y_1, and y_2, of

$$y_x = \frac{k}{1+e^{a+bx}},$$

let the problem be to find k, a, and b. First, plot the two points (y_0, y_1) and (y_1, y_2) on the nomograph, Chart 2. Then, draw a straight line connecting these two points and extend to intersect the 45° line, $y=x$. The x coordinate of this point of intersection is k. Connecting this point k with the point $(1+y_0, 1+y_0)$ and extending to the scale marked "$a+bx$" yields the value of a at the intersection. Similarly, connecting point k successively to points $(1+y_1, 1+y_1)$ and $(1+y_2, 1+y_2)$ yields $a+b$ and $a+2b$, respectively. The value of b is then determined by subtracting a from $a+b$; subtracting $a+b$ from $a+2b$ also gives a value of b and affords a check on the computations. Having determined a and b, we can now list the values of $a+bx$ for which y_x values are desired. The intersection with the 45° line of a line drawn between the point k on the y_x scale and $a+bx$ on the top scale yields $1+y_x$ on either the y_x scale or y_{x+1} scale from which y_x is readily determined by subtracting 1.

The mathematical solution of this problem in terms of y_0, y_1, and y_2 is

$$k = [2(1/y_1) - (1/y_0) - (1/y_2)]/[(1/y_1)^2 - (1/y_0)(1/y_2)]$$

$$e^{a+ib} = (k/y_i) - 1 \text{ where } i = 0, 1, 2, \text{ and}$$

where a and b are determinable from the last equation.

It can be shown[2] that if an equation $f(F, G, H)$ can be written as a determinant in the following form:

$$\begin{vmatrix} F_1 & F_2 & 1 \\ G_1 & G_2 & 1 \\ H_1 & H_2 & 1 \end{vmatrix} = 0$$

and three functions scales are drawn with coordinates $X_F=F_1$, $Y_F=F_2$, $X_G=G_1$, $Y_G=G_2$, $X_H=H_1$, $Y_H=H_2$ then the intersection of a straight line with these three scales yields a solution to the equation $f(F, G, H)$.

The above equation for k may be written in the following third order determinant:

$$\begin{vmatrix} 1/k & 1/k & 1 \\ 1/y_0 & 1/y_1 & 1 \\ 1/y_1 & 1/y_2 & 1 \end{vmatrix} = 0$$

which is the necessary and sufficient condition that the three points $(1/k, 1/k)$ $(1/y_0, 1/y_1)$ and $(1/y_1, 1/y_2)$ are collinear, when the function scales are determined as above. The above equation for e^{a+ib} may be written:

$$\begin{vmatrix} \delta/k & 0 & 1 \\ \delta/(1+y_x) & \mu/(1+y_x) & 1 \\ \delta e^{a+bx}/(1+e^{a+bx}) & \mu & 1 \end{vmatrix} = 0$$

which is the necessary and sufficient condition that the three points $(1/k, 0)$, $(1/1+y_x, 1/1+y_x)$ and $(e^{a+bx}/[1+e^{a+bx}], 1)$ are collinear, when the function scales are determined as before. Thus Chart 2 is basically double reciprocal paper and the scale at the top (from the above determinant) is $Y=\mu$ and $X=\delta e^{a+bx}/(1+e^{a+bx})$ where δ and μ are the total width and length of the chart grid, respectively.

The nomograph has five dotted lines showing the solution of the following specific problem: Given $y_0=2$, $y_1=4$, and $y_2=5$, to find k, a, b and some intermediate value of y_x, say, $y_{.5}$. First, the points (2, 4) and (4, 5) are plotted. The dotted line through these two points is extended to the line $y=x$. Then, reading the x coordinate of intersection on the y_x scale yields the value of k, namely, 5.3. The dotted line through the points (5.3, 0) and (3, 3) intersect the $a+bx$ scale at .51, which is the

[2] See Williamson, W. R. and Rasor, E. A., "Some Nomographic Theory and Applications to Benefits Under Retirement Plans," *Record of the American Institute of Actuaries*, XXX, 1941.

value of a. Likewise, the dotted line through the points (5.3, 0) and (5, 5) yields -1.10 for the value of $a+b$, and the dotted line through the points (5.3, 0) and (6, 6) yields -2.7 for the value of $a+2b$. By subtraction $(-1.10-.51)$, the value of b is -1.6, which is the same as

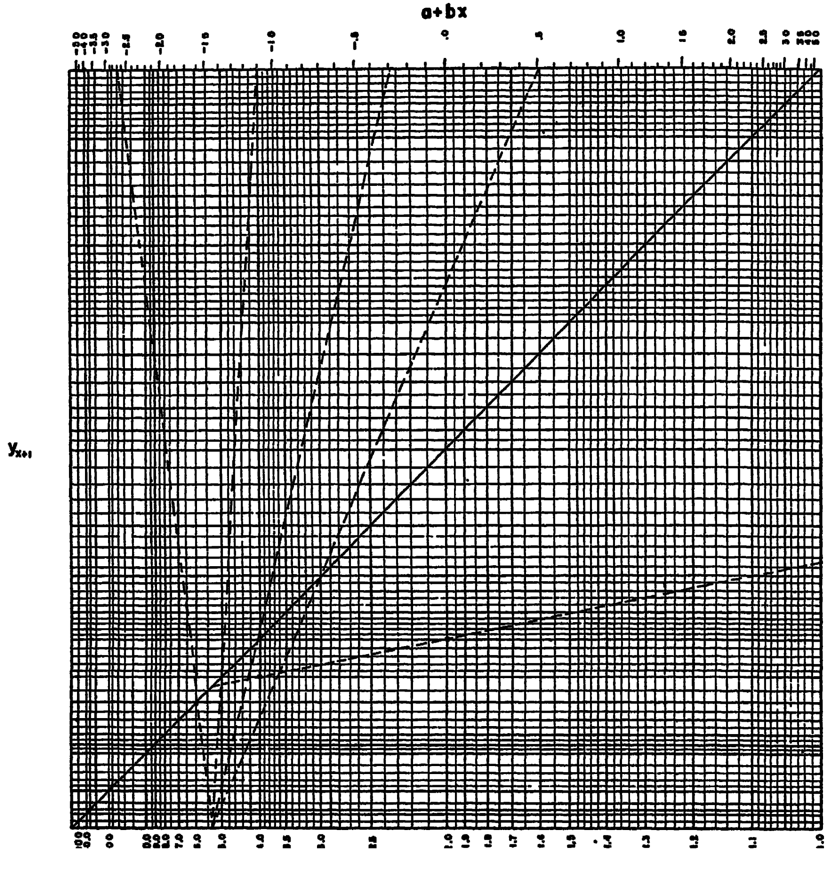


CHART 2. NOMOGRAPH FOR SOLUTION OF $y=\frac{k}{1+e^{a+bx}}$

obtained from the other subtraction, $-2.7-(-1.1)=-1.6$ so the computation is checked.

To find $y_{.5}$ we first determine $a+.5b$, which in this case is $.51+.5(-1.6)=-.29$ and the x or y coordinate of the intersection of the $45°$ line with the line drawn through 5.3 (the value of k) on the

y_x scale and $-.29$ on the $a+bx$ scale yields $1+y_5=4.1$ or $y_5=3.1$, the desired result.

In summary, it may be seen that the nomograph of Chart 2 has scales in simple linear form and may be entered directly without any preliminary calculations, while the result likewise is obtainable directly without any supplementary computations. A wide range is present for the various values of y, and if necessary this may be further broadened in three ways, by enlargement of the chart, a change of decimal place in all three given values of y, and by dividing the original data by y_0. These advantages indicate a much easier fitting of the logistic curve by a nomograph of the type of Chart 2, rather than one such as developed by Spurr and Arnold.

An even simpler nomograph for determining $1/k$ only would be to use a sheet of ordinary linear coordinate paper with a line drawn at 45° and the scales marked $1/y_x$ and $1/y_{x+1}$ in lieu of y_x and y_{x+1} shown in Chart 2. However, in using such a chart it would be necessary to predetermine the reciprocals of y and the result would yield the reciprocal of k. The scale for $a+bx$ could also be placed at the top as before, and resulting values for intermediate points would be $1/(y_x+1)$.

If more than three points are available, plotting all the points as above instead of a select three will show that the curve is of the logistic type if they lie on a straight line. On the other hand, if the points are not co-linear, an average value of k can be obtained from the chart.

ON THE BEST CHOICE OF SAMPLE SIZES FOR A t-TEST WHEN THE RATIO OF VARIANCES IS KNOWN

JOHN E. WALSH
The Rand Corporation

The situation considered is that of testing the difference of the means of two normal populations on the basis of a sample from each population, where the ratio of the population variances is known. The choice of sample sizes has been restricted to certain pairs which are equally preferable from the viewpoint of practical considerations (cost, difficulty of obtaining sample values, etc.). This note presents an easily applied method of determining which of these pairs of sample sizes yield the most powerful one-sided and symmetrical tests.

I. INTRODUCTION AND STATEMENT OF RESULTS

THE MOST powerful one-sided and symmetrical tests for comparing the difference of the means of two normal populations (variances unknown, ratio of variances known) with a given hypothetical value D_0 on the basis of sample values $x_1, \cdots, x_n$ from the first population and $y_1, \cdots, y_m$ from the second population are based on the t-statistic.

$$(1) \qquad \frac{(\bar{x} - \bar{y} - D_0)\sqrt{(n + m - 2)/(\theta/n + 1/m)}}{\sqrt{\frac{1}{\theta}\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}},$$

where θ equals the ratio of the variance of the first population to the variance of the second population (see [1]). This statistic has a t-distribution with $n+m-2$ degrees of freedom when the null hypothesis that the difference of means (mean of first population minus mean of second population) equals D_0 is true.

For the situation considered in this note, it is assumed that a number of pairs of sample sizes (n, m) have already been determined such that from the viewpoint of cost, inconvenience, etc., these pairs are equally preferable to the person applying the test; it is also assumed that the value of θ is known, either from past experience or by other means. The problem is to determine which of these given pairs of sample sizes yields the most powerful test (one-sided or symmetrical at the specified significance level) of the hypothesis considered. From above, the most powerful test will be based on the t-statistic (1) so

that this problem reduces to that of determining the pair of sample sizes which yield the most powerful test based on (1).

It is found (approximately) that the pair of sample sizes (n, m) which yields the most powerful one-sided test at significance level α and the most powerful symmetrical test at significance level 2α is that which furnishes the smallest value for the quantity

$$(\theta/n + 1/m)[1 - K_\alpha^2/2(n + m - 2)], \tag{2}$$

where K_α is the standardized normal deviate exceeded with probability α; i.e., K_α is defined by

$$\frac{1}{\sqrt{2\pi}}\int_{K_\alpha}^{\infty} e^{-x^2/2}dx$$

This criterion for choosing the pair of sample sizes which yields the most powerful test is reasonably accurate for $n+m \geqq 6$ if $\alpha=5\%$, $n+m \geqq 7$ if $\alpha=2.5\%$, $n+m \geqq 8$ if $\alpha=1\%$, $n+m \geqq 9$ if $\alpha=0.5\%$.

II. A FIELD OF APPLICATION

The results of this note are of most value when the ratio of the population variances can be considered known but the values of the variances are unknown. Some situations of this type are outlined below:

Let us consider a situation where the same treatment is applied to objects representative of populations of two different types. Then some common characteristic of the two types of objects is measured. The values thus obtained represent samples from each of the two types of populations, and these samples can be used to compare the relative effect of the treatment on the two types of objects (with respect to the specified characteristic). For example, the treatment might be a certain type of feed, the populations two different breeds of hogs, and the characteristic the weight at some specified future date.

Next let us consider the effect of different treatments on a fixed population. The mean and variance of the theoretical population of the values of the specified characteristic will vary with the treatment used. A not too uncommon situation is that where the variance of the theoretical population is a slowly increasing function of the mean of that population; i.e., if the magnitude of the observations increases, the variance of the observations also increases. Then a treatment which yields a theoretical population with a large mean value will yield a larger variance for this theoretical population than a treatment which results in a theoretical distribution with a smaller mean value. In the remainder of this section it will be assumed that the theoretical popu-

lations obtained for the specified characteristic have this property.

Finally let us consider applying a new treatment to two different populations of the type described above when data concerning the effect of some other treatment on these same populations are available. In many cases it can be assumed that the relative effect of the new treatment (with respect to the old one for which data are available) will be either to increase both of the theoretical population means or to decrease both of them. For example, an improved feed would be expected to increase the average weight of both breeds of hogs (though not necessarily the same amount) while a less satisfactory feed would be expected to decrease both average weights. If this is the case, the ratio of the variances of the theoretical distributions for the new treatment will tend to be about the same as the ratio of the corresponding variances for the old treatment; this ratio is therefore approximately known. Since the theoretical population means for the new treatment could be either larger or smaller than the corresponding means for the old treatment, however, estimation of the variances of the theoretical distribution for the new treatment from the corresponding values for the old treatment usually can not be done with any reasonable degree of accuracy. Thus the ratio of variances is available but the values of the variances are not.

III. EXAMPLE OF APPLICATION

Let us consider a case in which the only practical consideration is cost. Here the cost of an observation from the first population is always \$10, the cost of an observation from the second population is always \$100, while the total cost of the experiment is limited to \$400. Then the three pairs of sample sizes

$$(10, 3), \qquad (20, 2), \qquad (30, 1)$$

are equally preferable. Also for this particular situation it is known from past experience that $\theta=2$.

Let us determine which pair of sample sizes yields the most powerful symmetrical test at the 1% significance level. Then $\alpha=0.5\%$, $K_\alpha^2=6.62$ and the following table is obtained:

(n, m)	(2)
(10, 3)	0.76
(20, 2)	0.72
(30, 1)	1.20

Thus (20, 2) yields the most powerful symmetrical test at the 1% significance level.

A rough measure of the gain from using (20, 2) rather than (10, 3) is obtained by observing that (16, 2) furnishes approximately the same power as (10, 3). Thus about $40 is lost by using (10, 3).

If the significance level of the symmetrical test had been 10% rather than 1%, the following table would have been obtained:

(n, m)	(2)
(10, 3)	0.61
(20, 2)	0.64
(30, 1)	1.12

Then (10, 3) would yield the most powerful test rather than (20, 2).

Since for fixed n and m the power of a one-sided or symmetrical t-test based on (1) increases monotonely as the value of

$$\sigma_1^2/n + \sigma_2^2/m = \sigma_2^2(\theta/n + 1/m)$$

decreases (σ_1^2 = variance of first population, σ_2^2 = variance of second population), it might be thought that the pair of sample sizes which furnishes the smallest value of $\theta/n+1/m$ would yield the most powerful test. That this is not necessarily so is seen by considering the above examples. The criterion based on $\theta/n+1/m$ is independent of the significance level while the two examples show that the choice of the pair of sample sizes which yields the most powerful test varies with the significance level.

IV. DERIVATIONS

This section presents proof of the statement that to a reasonable approximation the pair of sample sizes which yields the most powerful one-sided test at significance level α and the most powerful symmetrical test at significance level 2α is that which furnishes the smallest value for (2).

Let μ be the mean of the first population while ν is the mean of the second population. First consider the one-sided test of $\mu-\nu<D_0$ at significance level α and based on samples of sizes n and m respectively. Using a modification of the normal approximation given in [2], it is found that the power function of this type (1) t-test is approximately equal to

$$N\left\{-K_\alpha + \frac{(D_0 - \mu + \nu)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}\,[1 - K_\alpha^2/2(n + m - 2)]^{1/2}\right\}$$

where by definition

$$N(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}dx.$$

This approximation is reasonably accurate for $n+m \geqq 6$ if $\alpha=5\%$, $n+m \geqq 7$ if $\alpha=2.5\%$, $n+m \geqq 8$ if $\alpha=1\%$, $n+m \geqq 9$ if $\alpha=0.5\%$.

Thus for $\mu-\nu<D_0$ and any permissible fixed values of α, σ_1^2, σ_2^2, D_0, μ, ν, the value of the power function is approximately largest when n and m are chosen so that (2) is as small as possible. This verifies the statement for the one-sided test of $\mu-\nu<D_0$. By symmetry the same result holds for the one-sided test of $\mu-\nu>D_0$.

Again using the modified normal approximation, the power function of the symmetrical type (1) t-test of $\mu-\nu\neq D_0$ is approximately equal to

$$2\alpha + \frac{1}{\sqrt{2\pi}}\int_{-K_\alpha}^{-K_\alpha+\delta} e^{-x^2/2}dx - \frac{1}{\sqrt{2\pi}}\int_{-K_\alpha-\delta}^{-K_\alpha} e^{-x^2/2}dx, \tag{3}$$

where

$$\delta = \frac{|D_0 - \mu + \nu|}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}\left[1 - K_\alpha^2/2(n + m - 2)\right]^{1/2}.$$

Since $2\alpha<1$, $\alpha<0.5$ and (3) is a monotonely increasing function of δ. Thus for any fixed values of α, σ_1^2, σ_2^2, D_0, μ, ν the value of the power function is approximately greatest when n and m are chosen so that (2) is minimum. This verifies the statement for symmetrical tests.

REFERENCES

[1] Henry Scheffé, "On solutions of the Behrens-Fisher problem based on the t-distribution," *Annals of Math. Stat.*, Vol. 14 (1943), p. 43.

[2] N. L. Johnson and B. L. Welch, "Applications of the non-central t-distribution," *Biometrika*, Vol. 31 (1940), p. 376.

NOTE ON SOME ERRORS IN "THE EVIDENCE OF PERIODICITY IN SHORT TIME SERIES"

Several fundamental errors are in Truman Kelley's "The Evidence of Periodicity in Time Series." These involve a confusion of numbers of observations and degrees of freedom, the interpretation of probabilities obtained in tests of hypotheses and the appropriateness of periodigram analysis for detecting the existence of periodicity in time series.

I

THE ARTICLE by Truman L. Kelley, "The Evidence of Periodicity in Short Time Series"[1] contains several fundamental errors which should be noted. These particular errors are fairly widespread in popular literature and in practical applications. If these errors occur in this Journal and are not corrected their propagation will be insured since Kelley is deservedly a leading authority in Statistics. The younger generation, as well as the older, frequently rely on authority.

Originally, in 1943, the writer had considered it not worthwhile to write this corrective note. But recently on two occasions he has seen others using the technique presented by Kelley. On several other occasions during the war, interpretative errors to be mentioned below were made by persons who had taken formal university statistics courses and who were regular users of statistical methods. It is believed worthwhile, therefore, to point out these fundamental errors.

1. In testing goodness of fit of a trend line to a time series, Kelley treats the number of observations as independent observations in computing the number of degrees of freedom.[2] It is obvious (practically) that the observations are not independent.

2. In estimating and testing the period of a series of residuals, after removal of trend, the number of observations is again used in order to determine the number of degrees of freedom. In this particular analysis it might be contended that if periodicity does not exist in the residuals the residuals are independent; and hence, the assumption of independence is acceptable for purposes of testing the null hypothesis of non-periodicity. This contention would overlook the fact that the

[1] Truman L. Kelley, "The Evidence of Periodicity in Short Time Series," *Journal of the American Statistical Society* (1943), Vol. 43, pp. 319–326.

[2] *Ibid.*, pp. 320–325.

H T. Davis, *The Analysis of Economic Time Series* (Bloomington, Indiana, 1941).

H. Wold, *A Study In the Analysis of Stationary Time Series* (Uppsala, Sweden, 1938).

Maurice G. Kendall, "On the Analysis of Oscillatory Time Series," *Journal of the Royal Statistical Society*, Vol. CVIII (1945), p. 93. This more recent article is cited because of its general excellence and comprehensiveness.

power of the test of non-periodicity as against various alternative lengths of periods is a function of the number of degrees of freedom (ignoring the extension into the sphere of varying amplitudes). As alternative hypotheses about the length of the period are admitted, the number of degrees of freedom applicable changes. In particular, the longer the period the fewer the degrees of freedom. If the degrees of freedom were constant, the test would have a greater probability of rejecting the non-periodicity hypothesis as the admissible alternative hypotheses of other period lengths are broadened to include longer periods. However, the degrees of freedom *do* decrease so that on balance if one assumed a fixed number of degrees of freedom he would not be controlling the probability of errors of the first type—namely rejecting the non-periodicity hypothesis when the series is in fact non-periodic.[3]

3. A common error of interpretation is present also. "Values of progressively less and less than .5 provide greater and greater evidence that b is not a chance deviation from zero."[4] By itself this statement is correct if it means that small values of P make it harder to accept the hypothesis that b is equal to zero (for a given power function). But it is apparent that other connotations are implied. "As P from this variance ratio is less and less than .5, there is greater and greater evidence that a period in the neighborhood of T exists."[5] If the test is biased there is no such evidence. The degree to which evidence is discriminatory is dependent upon the power function. In simple terms, evidence favors one alternative hypothesis only if it is relatively more probable under it and relatively less probable under other hypotheses. The following statement appears a little later: " . . . $P=.048$. Having odds of 952 to 48 that -1877.9 is not a chance deviation from zero. . . . "[6] The second sentence is a *non-sequitor*. If it were a valid extension, it would mean presumably that of all the times that one observes such deviations (those giving $P=.048$) .952 of such cases would in the long run have occurred from parameters not zero. Only by using Bayes' theorem can one obtain the type of statement employed by Kelley. The proportion of times in which such deviations are observed is a function of the experienced (true) parameters and not at all a result of the particular observed sample in any given case. What the probability does mean is that the probability of getting such a chance deviation or worse, *if* there is non-periodicity, is equal to .048. One *cannot* simply interchange

[3] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Phil. Trans. Royal Society of London*, Series A, Vol. 231, p. 289.

[4] Kelley, *supra.*, p. 320.

[5] *Ibid.*, p. 323.

[6] *Ibid.*, p. 325.

the role of the hypothesis and the sample evidence and still have a correct statement.[7]

Later there appears "The smallest P is .152, found when $T=16$. Thus the odds are about 5 to 1 that a period in the neighborhood of 16 years exists."[8] In this statement and procedure of getting a P of .152 several errors are combined. a) The selection of the smallest probability from a set of independent probabilities will result in obtaining too many (more than would be obtained by a random selection) significant probabilities. This point has been covered by others.[9] It is especially applicable to periodigram analysis. b) The probabilities from the periodigram are not independent so that it is impossible to apply correct methods designed to test the significance of the largest (or smallest) of a set of independent probabilities. c) Finally the second sentence is again a *non-sequitor*. Suppose the probability had been .001. Would that then indicate that the odds are 999 to 1 that this is not a chance deviation and is instead a real difference? Obviously not. It merely says that *if* the truth is non-periodicity, then one would obtain by chance such a deviation only once in 1000 times (more accurately with probability .001). One cannot interchange the role of hypothesis and sample evidence! Consider the other extreme probability; suppose the probability in the periodigram for a particular period is 1.00. Clearly that is not *proof* that such a period is not in the series! Not only is the second sentence a *non-sequitor*, it is wrong. If it were correct it would be necessary that of all the cases in which deviations with probability .15 were observed, .85 of *those* cases would ('on the average') have come from situations in which there really was periodicity of length $T=16$! This requires Bayes' theorem. There is no indication that Kelley has anywhere in mind any *a priori* parameter distribution. And if he did, the simple complement of the probability is not the correct one if one wishes to apply Bayes' theorem.[10] The appropriate formula is somewhat more complex.

In summary of this particular error, the confusion appears to arise basically from a failure to understand adequately the fundamental procedure by which one increases his knowledge on the basis of evidence. Always one must ask how probable is the evidence that has been observed, *if* a certain hypothesis or cause is the true one. He *must also* ask

---

[7] Neyman and Pearson, *supra*.

[8] Kelley, *supra*, p 326.

[9] W. G Cochran, "The Distribution of the Largest of a Set of Variances as a Fraction of Their Total," *Annals of Eugenics* (London), Vol 11, p. 47.

[10] J V. Uspensky, *Introduction to Mathematical Probability* (New York: McGraw-Hill, 1937), pp 60–73

the same question about the probability of the observed evidence under the assumption that it came from some *other* hypotheses. After comparing these probabilities of obtaining the evidence (not of the hypotheses!) he can venture to draw conclusions. Never can he draw a conclusion until he has so obtained these different probabilities. At worst, he must make some intuitive guesses about the relative probabilities under various possible hypotheses. Lacking even an intuitive guess he can draw absolutely no conclusion whatsoever! And unless one uses Bayes' theorem, he simply cannot proceed to the next desirable (but nevertheless impossible step) of making a statement about the probabilities of the hypotheses![11]

4. Finally, the basic procedure used to detect periodicity—the use of a periodigram—is inappropriate. The periodigram technique used is appropriate to determination of the length of the period if the series is known to contain a period. Kelley's problem however is that of determining whether or not one exists.[12]

The writer regrets the occasion for this note, but these same errors are altogether too frequently committed elsewhere. It is another example of the adage that a statistician probably spends more of his time telling what not to do rather than doing.

ARMEN A. ALCHIAN

UNIVERSITY OF CALIFORNIA, LOS ANGELES

II

I do consider the number of observations as the number of independent items of information, or the number of degrees of freedom. No matter how brief the time interval between observations, if they are not consequent to the same instrumental errors—including judgment errors if estimates have been made—I would say that there is justification for calling them independent measures. If there is some other standard I fail to grasp it. However, I do not see that this point is crucial to Alchian's further criticisms.

Before discussing the main issue, which is that of inverse probability, I will comment upon the criticism of my statement that "The smallest P is .152. . . . Thus the odds are about 5 to 1. . . . " The value $P=.152$ was taken from a table giving P values for 17 different periods, 2-years, 3-years, . . . 18-years. As the value in that table there is no act of selection, but the moment it is taken out of that table because it is the

[11] Neyman and Pearson, *supra.*

[12] M. G. Kendall, *supra.*

smallest and put in the text there is an act of selection, so I admit that my statement, correct in the table, is a misstatement in the text and that Alchian's point (a) is sound. I am not sure about his point (b), but if in place of "impossible to apply" he had written "impossible for Kelley to apply" I would fully subscribe to it. His point (c) is in part covered in the first of this paragraph and is otherwise approached in the following discussion of inverse probability.

In the universe of situations in which designated null hypotheses do not hold other null hypotheses will hold. One very large class of these will be those in which the shift of a single parameter will lead to a null hypothesis which now holds. True, this is just a sub-class of the universe of situations, but a sufficiently large sub-class that a statement of probability covering it will be useful in determining conduct. In fact it seems to the writer that for most, perhaps all, of the problems of life a more valid determiner of conduct is unavailable. To illustrate: We have a null hypothesis that the true mean is M_1. The actual sample mean is M and we find that $P=.048$. That is, if the true mean is M_1 a deviation as great as that found will, in the long run, occur in .048 of the samplings. Now consider the divergence from the sample mean M of the true mean $\tilde{M}$. All parameters except the mean being constant in many samples of different sorts all of which have a sample mean M the true means will diverge from M by an amount greater than (M_1-M) in .048 of the samples. Thus in this restricted realm inverse probability holds. (We may note in passing that this restricted realm may be far less restricted by making linear transformations of the original variable.)

Let us say that this probability is small enough to lead to a certain course of action. Nothing but time and further experience will prove whether the course of action is sound or not. A double risk has been involved: first the $P=.048$ instead of some smaller amount and second is the materiality of other parameters than the one studied. As a matter of logic both of these risks must be present.

To set up alternative hypotheses and determine which of the two is more acceptable does not change the situation. It sharpens the issue so far as P is concerned, but it may augment the other risk for we must ask if the two hypotheses were the best two to use in determining future action.

Though I subscribe to the soundness of Alchian's reasoning I hold that the use of inverse probability has much practical warrant and usefulness. I suspect that in deciding upon a course of action from given

data Alchian and I would reach the same conclusion, though we employ somewhat different logical processes en route.

Alchian's final point—that the periodogram cannot be used to detect periodicity—is indeed sweeping. I know of no logical approach to the question of existence independent of that of amount to which the thing exists. Dr. E. L. Thorndike's oft quoted dictum "If a thing exists it exists in some amount" has been most serviceable in the field of psychology. It should equally serve other fields.

I appreciate the fine tone and precision of Alchian's article and have the hope that we are really not as far apart as his criticisms imply.

TRUMAN L. KELLEY

WILLIAM LANE AUSTIN (1871–1949)
JAMES CLYDE CAPT (1888–1949)

Two former Directors of the Bureau of the Census, both valued members of the American Statistical Association, have passed away in recent months. William Lane Austin, Director of the Census from 1933 to 1941 and a Senior member of the Association, died in Greenville, Mississippi on October 10. James Clyde Capt, who succeeded Austin as Census Director in 1941 and remained Director until his resignation because of illness on August 10, 1949, died in Washington on August 30.

A review of the careers of these two men—alike in some respects and differing widely in others—brings into sharp focus some of the changes during recent decades in our concepts of the role of statisticians in government service. Austin entered the temporary Census Office in 1900, two years before the Permanent Census Act created the present Bureau. He served successively as statistician in charge of the Census of Plantations; Chief Clerk; chief statistician in charge of the Censuses of Agriculture in 1920, 1925 and 1930; Assistant Director; and finally Director.

During this 40 year period he learned his statistics from experience with Census operations. And despite the brilliant analyses of census problems by General Francis A. Walker and some other former Directors, it was not generally recognized that Census operations required special skill or presented technical problems. Austin developed hard-headed shrewdness and skill as a negotiator and as the supervisor of personnel whose training in most cases, like his own, had been limited to experience in the ranks. It was when he became Director in 1933 that he first displayed abilities and attitudes that co-workers of a lifetime to quote the late Joseph H. Hill "had never suspected him to possess."

The immaturity that had previously characterized statistics as a profession was beginning to disappear and Austin felt the need to strengthen the personnel of the Bureau. With astonishing finesse in avoiding offense to life-long colleagues in the service—men who distrusted "impractical theorists"—he began building a staff having technical qualifications that he did not himself possess or even understand. Austin was the founder of the modern, efficient, and technically competent Bureau of the Census that serves the nation today.

Capt entered the Bureau of the Census in 1939, as Executive Assistant to Austin. Unlike the latter, he was without any previous experience in dealing with technical statistical problems, and was completely unknown to the statistical fraternity and to those users of data throughout the land who view the Census Bureau with somewhat the

same veneration that lawyers hold toward the Supreme Court. It is no exaggeration to say that many of these were profoundly apprehensive two years later when the still new Assistant was named Director, at Austin's retirement.

I well remember my own doubts en route to my first call upon the new Director, and the sense of reassurance with which I left his office. Nor was the spirit of trust and understanding engendered at that first contact ever dimmed in any of our later official or personal relations. Increasingly I came to regard "J.C.," as he was affectionately known, as a superb administrator. His direction of the Bureau's affairs impressed me as strict, unaffected by "old school ties," personal friendships or biases, and always fair and just. He was quick to admit mistakes and shortcomings, but coupled his admissions with determination that the same mistakes would not occur a second time. In his relations as Director of the Census with other agencies, he was always cooperative, unambiguous in his position, and decisive.

I believe, however, that the achievements of his administration sprang primarily from the quality that he shared so conspicuously with Austin, namely: respect for and deference to competence in others. This seems to me to have gone farther than the layman's frequent deferral to the specialist, and to have included a willingness to give his personnel free rein, even in areas of administration in which he could have claimed a superior right to judgment. They must prove their initiative by success, but the opportunity thus presented them reflected an inherent and commendable modesty in their chief. It made for staff loyalty.

The administrations of these two Directors represent, very fortunately, a consistent and continuous pattern of growth. The strength now exhibited by the Bureau of the Census, particularly in the high calibre of its technical and administrative staff, had its beginnings under Austin, and developed at an accelerated pace under Capt. The all-important need of the Bureau was the building up of the organization to a position of leadership in statistical techniques and operating methods. This seems to me the outstanding and culminating achievement of Capt's administration. Credit for his success must be shared with Philip M. Hauser, Ross Eckler, Howard Grieves and his other top assistants, upon whom the burdens of the Bureau administration during the crucial period of the Seventeenth Decennial Census now fall. Their tasks will be easier because of the solid process of Bureau-building that preceded.

STUART A. RICE
Assistant Director in
Charge of Statistical Standards
Bureau of the Budget

BOOK REVIEWS

(Dr. Oscar Buros' resignation as Review Editor was effective August 1, 1949. During the interim period until a new editorial committee is established, this section will be edited by Dr. Ernest Rubin for the Secretary's Office.)

Probability Theory for Statistical Methods. *F. N. David* (Lecturer in Statistics, University College, London, England). London and New York: Cambridge University Press, 1949. Pp. ix, 230. 15*s*.; $3.50.

Review by John W. Tukey
Associate Professor of Mathematics
Princeton University, Princeton, New Jersey

Sixteen chapters of from 11 to 18 pages each, and a short seventeenth, emphasize that this book is based on a carefully organized set of lectures. The imprints of Karl Pearson and Jerzy Neyman are plain to see. On the one hand, the introduction of the binomial distribution, in Chapter 3, is followed by derivations and applications of the evaluations of binomial probabilities with the Incomplete Beta Function Table, Uspensky's method using hypergeometric series, the normal approximation, and the Poisson distribution (which is regarded solely as a limit to the binomial). On the other hand, the Markoff theorem on least squares is carefully applied to stratified samples of minimum variance.

Noteworthy are (*i*) a 17-page chapter on simple genetical applications, (*ii*) the elementary derivation of sampling moments for sample moments from finite populations, (*iii*) an account of differences of zero and their simplest uses, (*iv*) an emphasis on mathematical precision and the use of elementary methods, and (*v*) a clear-headed middle-of-the-road attitude toward the relation of probability theory to the real world. While it is hard to name courses in the United States where such a book would be an appropriate text, it should be helpful supplementary reading for both students and teachers. (The reader is supposed to be familiar with the notation for third and fourth moments (p. 31).)

As is inevitable and appropriate, there are places where the author and the reviewer hold different views, but in only two places were definite errors detected. One is in the statement of the boundedness hypothesis for the Central Limit Theorem on page 217, where $0 \leqq m_2$ should read $0 < m_2$. The second is the derivation of Sheppard's corrections on page 214, where the cumulants of G are expressed in terms of those for E and x as if E and x were independent, when in fact E determines x completely.

The reviewer is unable to be sure of the meaning of "The fundamental probability set, written F.P.S. for short, will be just that set of individuals or units from which the probabilities are calculated" (p. 12) in view of the discussion on page 58. The only other place which seems likely to confuse

the student is in the discussion of the "elementary probability law" of a continuous random variable, where a careless reader could conclude that $p(x')$ was "the probability that x takes the value x'."

The author is definite on such points as replacing the binomial distribution by a continuous distribution with stepped densities, "Such an assumption is, of course, wholly fallacious, since the binomial probabilities are a discrete set of points, but it is a useful aid to memory if not pursued too tenaciously" (p. 53—more tenacity would have replaced npq by $npq+(1/12)$ with better approximation), the uselessness of the negative binomial, "It would appear wrong therefore to carry out calculations in which p and n are given negative values" (p. 65) and division by n or $n-1$, "If a measure of the scatter in the sample is required then—(dividing by n)—must be calculated. If it is desired to estimate the population standard deviation then the expression—(dividing by $n-1$)—may be calculated because in the long run it will be equal to σ" (p. 126). Two points need to be made in the last connection. First, if "in the long run" were well defined, the last statement would be wrong. Second, the reviewer, and, he believes, an increasing number of other statisticians, feel the time is coming soon to loosen the thrall of moments of inertia, and *always* divide by $n-1$.

References to further reading are given after each chapter. These could be slightly improved by giving dates (only 4 are given, including one on page 81) and by giving volume numbers in Arabic rather than Roman numerals (after all, statisticians no longer do arithmetic in Roman numerals).

After mentioning a few of the things he likes, and all of those he doesn't, it behooves the reviewer to point out what he failed to find. To fill the two prominent gaps in a reasonable manner, Miss David would have had to recast her order of presentation and bring in cumulants (semi-invariants) long before the next-to-last-chapter. If this had been done, and k-statistics had been added, then Irwin and Kendall's (*Ann. Eug.* 12: 138–142, 1944) powerful and simple methods of deriving sampling moments of sample moments from finite populations could have been used with a great gain in simplicity of presentation and in power of tools. (The experience of the early workers with heavy algebra and consequent errors points up the need for keen and powerful tools.) Second, the early introduction of cumulants would have made it possible to extend the present discussion of Lexis theory, which seems ineffective, into a presentation of the essentials of the analysis of variance *without normality assumptions* as developed by Pitman and Welch (cf. *Biometrika*, Vol. 29). This would have been a most important addition.

Quality Control by Statistical Methods. *G. Herdan* (Lecturer in Statistics, Department of Preventive Medicine, University of Bristol, Bristol, England). Edinburgh, Scotland: Thomas Nelson & Sons, Ltd., 1948. Pp. xi, 251. 21*s*.

REVIEW BY PAUL PEACH
Associate Professor, Institute of Statistics
University of North Carolina, Raleigh, N. C.

THE internal evidence of Dr. Herdan's book proves that he has at least some acquaintance with a wide variety of statistical methods. In his eight chapters he discusses not only the usual material of quality control literature (charts for fraction defective, charts for variables, and single and double sampling plans based on attributes) but such less customary subjects as correlation, analysis of variance, and inverse probability. He devotes a whole chapter to the use of probability graph paper, a tool that certainly deserves to be introduced to industrial workers. His is, I believe, the first text on statistical quality control to mention the important notion of components of variance. He includes notes on the t and chi-square tests, with a table of t (but not χ^2, "in order not to swell the number of appendices inordinately"). He has a brief note about generating functions. There is perhaps not another book in the statistical field that touches upon so many topics in so little space.

Elementary books are presumably written for the instruction of students, the general idea being that beginners are expected to read the book and learn from it certain lore that shall be true, or useful, or at any rate acceptable to authority. The student who picks up Dr. Herdan's book with any such expectation had better be prepared to take a substantial consumer's risk. Naturally, no ordinary book of 250 pages can include adequate expositions of all the topics Dr. Herdan has introduced. I don't believe anybody will learn to fit the least squares straight line by studying pages 115–118. The Latin square on page 102 is almost void of explanation or context. The discussion of acceptance sampling is based mostly on the work of Dodge and Romig; but there is only half a sentence about the AOQL concept. The professional statistician learns from these pages that Dr. Herdan knows about linear regression and Latin squares; the beginner can learn, I fear, nothing.

The book suffers further from a neglect of common standards of literary craftsmanship. Both the writing and the typography are frequently obscure and confusing, and sometimes we find downright misstatements that are obviously the result, not of ignorance, but of carelessness. On pages 3 and 4 we find what purports to be the equation of the normal curve, but so carelessly set up that no novice could be expected to read it rightly. In the discussion of the test for the significance of the difference of two means, some of the formulas are right, some wrong. In the chapter on probability graph paper the probability density function and the ogive are persistently confounded.

Even in his literary references Dr. Harden fumbles. He uses for one chap-

ter the whimsical subtitle "The Rod of Moses"; the rod in question was Aaron's (Exodus 7: 8–12) and the allusion seems hard straining for a defective analogy. Among his American authorities he mentions one, called Cowder in one place and Crowder in another; the reference is to my esteemed colleague, Prof. Dudley J. Cowden. This latter error is in some respects typical. It could have been avoided with a minimum of care; people who know Dr. Cowden will penetrate the veil of obscurity; others will derive no information from the reference.

One wonders how Thomas Nelson & Sons came to publish this book, what authorities served as their referees, whether the comments of these authorities were transmitted to Dr. Herdan, and what was done with the galley and page proofs. Something must have been neglected somewhere. At all events, the resulting book has in my opinion no possibilities for the instruction of students.

Statistical Methods in Research. *Palmer O. Johnson* (Professor of Education, University of Minnesota, Minneapolis, Minn.). New York: Prentice-Hall, Inc. (70 Fifth Ave.) 1949. Pp. xviii, 377. $7.65.

REVIEW BY FREDERICK MOSTELLER

Associate Professor of Mathematical Statistics, Department of Social Relations Harvard University, Cambridge 38, Massachusetts

A GREAT many people have probably been intending to write a modern statistical textbook similar to Johnson's. The notion is to write a rather advanced book suitable for students in education and psychology, suitable in the sense that the emphasis shall not be on the mathematics, but upon methods which are useful in research problems the students will face. People intending to write such a book now have two alternatives: they can either relax and forget about it, or they can raise their sights a good deal higher than they needed to before the publication of Johnson's book.

The arrangement of the material resembles that of a handbook, which is rather appropriate in view of the title. This arrangement will facilitate use of this book by active research people who will find the form very convenient. The table of contents is helpful in this connection. The index looks impressive, but the only thing I looked for I was unable to find (see below). The methods used are described carefully, but succinctly, and illustrated by worked-out problems on real data. Methods given are heavily referenced and every effort is made to send the student to original sources. I do not have much hope that students will go to the sources, but do believe that teachers will be grateful for the references because not all who will want to use this book will be prepared to teach from it. This book can be enhanced by instructors able to expand on the material presented. The first 200 pages could easily be inflated to 400 with no padding.

Tables of the normal (a poor one), t, chi-square, F, and one of Nayer's

for testing differences in variance among several samples of the same size are supplied. Over the years one gets sadder about tables of the normal. Why is it that authors only provide half of one table? It is perfectly true that by a little arithmetic one can get what one wants from such a table, but tables are for economy in time and errors. As a start I would like several normal tables: 1) cumulated from the left, 2) cumulated from the right (these should include negative arguments as well as positive), 3) cumulated symmetrically from the center, 4) cumulated symmetrically from the tails. I would also like companion tables using probability as the argument and the deviation as the entry. If it is objected that this is pure laziness, the energetic may still do their computations from the original function. I hope some textbook writers and publishers will cooperate in this matter soon.

The book opens with a much better than usual discussion of the realm of statistics. To the discussion of general uses of statistics in economic research I would add the possibility of statistical or probability models. In the chapter on probability and likelihood the author is not especially careful about distinguishing between the notions of a probability limit (p. 20) and the usual concept of limit. Bayes Theorem and Maximum Likelihood are too briefly treated, but the latter is discussed more fully later. The book begins to open up in Chapter III on Sampling Distributions, discussing means, differences between means, variance, t, r, the z transformation for correlations, the relation between z and F. Discussion of testing statistical hypotheses follows smoothly, comparing fiducial and confidence limits (making confidence limits sound a little more difficult than necessary), likelihood ratio and sequential tests.

Standard procedures are given for testing means against standards with standard deviation known and unknown, for finite as well as infinite populations, testing differences, the Behrens-Fisher problem, the sign test, testing differences in percentages using t. There is a footnote on page 81 which bothers me in this connection—"The χ^2 test is an exact test for this problem." If no correction is made for continuity the stated formula is equivalent to χ^2, but neither test can be regarded as exact as I see the matter. Continuing with this chapter, methods are given for testing equal variability for two or more groups, the significance of differences for correlations and regression coefficients, 2×2 and larger tables, goodness of fit. I will not continue this parade of techniques, but mention that later chapters cover estimation, interval and point, normalized distributions, special devices when the data are nonnormal, sampling theory and practice. All this is in 200 pages. The second half of the book handles analysis of variance and covariance, and multiple regression, including the discriminant function.

The analysis of variance examples are carried out in detail, one example running through 14 pages. The discussion is better than most such. Not enough emphasis is put on the meaning and interpretation of interactions, not enough is said about the limitations of the methods. I doubt that educators are always interested in such questions as: Are the mean achievements

of students in the three grades equal, or are the school means on the reading test equal? This is not so much a shortcoming of the book, but a shortcoming of the method as it is now available. This author does give with no heading, and no reference in the index that I can find, Fisher's method for comparing two particular means selected from a set (p. 234).

Perhaps the best thing to say is that Prentice-Hall's advertisement for this book is correct in all particulars—except one, I doubt if the author did research with R. A. Fisher (sic), although the influence of R. A. is clear.

Psychological Statistics. *Quinn McNemar* (Professor of Psychology, Statistics, and Education, Stanford University, Stanford, California). New York: John Wiley & Sons, Inc. (440 Fourth Ave.) 1949. Pp. 364. $4.50.

REVIEW BY EDMUND CHURCHILL

Assistant Professor of Mathematics, Department of Mathematics Antioch College, Yellow Springs, Ohio

PROFESSOR McNemar's book is described as covering "all the statistical techniques, except factor analysis, that are frequently useful in psychological research." Altho "frequently used" might be more accurate than "frequently useful," and there are omissions in either case, this book does cover the standard topics from descriptive statistics through correlation, confidence limits, chi-square, sampling, and the analysis of variance, plus a chapter on the too often neglected topic of analysis of covariance.

The book, unfortunately, falls short of the author's aim to "provide a concisely, yet clearly written textbook which will lead to an appreciation of the place of statistics in psychological research." Conciseness is far from apparent at many places in the book, and neither the writing nor the logic behind it are always clear. This reviewer believes strongly that the student will best gain an appreciation of the role of statistics if clear, meaningful, and interesting illustrations are given as each technic is introduced. In this respect this book is decidedly weak. At least a dozen of the technics or procedures discussed in the book are not illustrated at all. Some are illustrated by ficticious data, others by data of the type: two groups classified according to 5 response categories (what groups? what categories?), 13 pairs of scores, IQ's of 161 five-year old boys, scores made by 50 college men on the Brown spool packer, etc. Here was an excellent opportunity both to enliven the text and to point the way to sound research practices by the frequent use of data from well designed (and well described) research.

There is a fuzziness in the discussion at a number of points. The discussion of the null hypothesis (p. 223) as it relates to the difference between means is not only muddled but at variance with an earlier discussion. The hypothesis of equal means does *not* imply equality of the population variances nor does the hypothesis of equal means and variances lead to the t test. It is

interesting to note that, at different places in the book, the same hypothesis of equal means does not involve equality of variances (p. 65), is said to imply equality of variances (p. 223), and is described as involving such equality as an assumption (p. 225). The author would have stayed within his purposes if he had not only avoided this confusion of hypotheses and assumptions, but had also included tests of the hypothesis of equal means and variances and of the hypothesis of equal means without the assumption of equal variances.

The spirit of the null hypothesis is well violated in a procedure for testing the quantity $(p_1-q_1)-(p_2-q_2)$ where the p_i and q_i are non-exhaustive proportions in independent samples. Standard errors are obtained for (p_i-q_i) on the basis of the null hypothesis: $p_i-q_i=0$ $(i=1, 2)$. These standard errors form the basis for testing the difference mentioned above, altho the hypothesis that this difference is zero is a far cry from that used in getting the standard errors. The concept of the quality of a statistical test or confidence limit is missing. Obviously, no technical treatment of this concept is feasible but the awareness that we use one test or another in a given situation because we believe it is less likely to lead to error and that no test is always the best one is surely a basic part of the understanding of statistical inference. There are a number of minor points about the content of the book which may be worth noting. The square root of s^2 is, as usual, incorrectly described as unbiased. The sequence of histograms of $(\frac{1}{2}+\frac{1}{2})^n$ with fixed base length is asserted to converge to the normal curve; the sequence actually converges to a straight line. Sheppard's and Yate's corrections are given without hint that they can worsen instead of improve matters; this is especially true when Yate's correction is applied, as McNemar recommends, to 2×1 tables. The labelling of McNemar's graphs and the stubs of his tables, are, in several cases, atrocious and his treatment of approximate data is on the same level. There is no hint that in stratified sampling we might use samples in which the proportions are not in the same ratio as in the population, altho in general the best design calls for different ratios in sample and population. The equating of stratified sampling with the quota method may well confuse the student who has understood the quota method to be the one used by Gallup.

Frequent use of elementary algebra is made to derive formulas or to point the direction of their derivation. Altho there are those who will object to this practice, it is undoubtedly sound and a practice to be encouraged. On the other hand, it is difficult to understand why calculus is introduced to half-derive the linear regression equations when the elementary algebraic derivations of these equations are so simple and straightforward. Incidentally, a little of this algebra applied to the chi-square formula on page 207 would show that the maximum value of C for $2\times n$ tables is $\sqrt{1/2}$ and hardly rates classification as "unknown" (p. 182). One can also demonstrate that the maximum value of C for any $m\times n$ table is the same as for an $m\times m$ or an $n\times n$ table, whichever is the smaller.

The discussion of test reliability presents clearly some of the difficulties in estimating reliability, but its advice to base estimates on parallel forms is often completely unrealistic. The construction and administration of a second test form, if possible, would often represent an extremely inefficient use of research facilities. Split-half methods are mentioned but nothing is said about the effect of speed on these methods. The useful, reasonably sound, and rapidly computed Kuder-Richardson no. 20 is unmentioned. The simple derivation of this formula by Jackson and Ferguson might well have been included here.

The importance of homoscedasticity is stressed thruout the section on correlation (and might well have been stressed more often in the section on analysis of variance) but McNemar gives no hint as to how one may test a set of data for this mouth-filling property. Nor is the warning that many of the technics discussed are applicable only to normal data accompanied by advice or reference to sources of advice for the non-normal case. Much of this reviewer's pleasure at finding a chapter on covariance analysis is offset by the omission of such things as the discriminant function and the equivalent T^2 test, more useful, to be sure, than used, and the growing body of rapidly computed "inefficient" statistics which can be of considerable value in the early stages of a research program.

The above remarks are not intended to imply that there is not much which is clear and sound in this book. But statistics has reached a point in its development where we can rightfully expect that beginning texts have the clarity and thoro soundness that we expect and find in a college algebra text and do not find in this text.

Quality Control in Production: A Machine-Shop Manual on the Statistical Method of Controlling Product Quality During Manufacture. ***H. Rissek.*** Foreword by Frank Gill. London: Sir Isaac Pitman & Sons, Ltd., 1947. Pp. xvii, 181. 21*s*.

REVIEW BY H. A. FREEMAN

Associate Professor of Statistics, Massachusetts Institute of Technology Cambridge, Massachusetts

THE title and subtitle describe the purpose of this book. Its chapter headings indicate its content. They are: 1, Statistical Method and the Quality Problem; 2, What is Quality Control? What are its Advantages? 3, The Basic Principles of Quality Control; 4, Control Charts based on Dimensional Measurement; 5, Control Charts based on Counting Defectives; 6, The Organization of a Quality Control System. There are also an adequate set of tables, a fair index, a good bibliography, and a foreword by Sir Frank Gill, Past President of the Institution of Electrical Engineers.

It is hard for me to say if this book is better or worse than its many competitors. Its statistical level corresponds to that now standard among qual-

ity control manuals; it must therefore be rated on its effectiveness in educating factory personnel. On this, I can only guess; my guess would be that this is a superior book, neither as facile or extended as some American manuals, but one which shows both that the author has digested his experience in industrial quality control and that he knows how to write about it.

One novelty: A discussion of Tippett's dual control chart is included. In this plan, the sampling results of go and of no-go gauges are used separately, the difference between the two numbers defective being a good estimate of the dimensional average, the sum, a good estimate of dimensional variability.

Sampling Methods in Forestry and Range Management, Second Edition
F. X. Schumacher (Professor of Forestry, Duke University, Durham, N. C.) and *R. A. Chapman* (United States Forest Service, Washington, D. C.). Duke University School of Forestry, Bulletin 7, Revised. Durham, N. C.: the School, June 1948. Pp. 222. Paper, $2.00; cloth, $2.50.

REVIEW BY WALTER H. MEYER
Professor of Forest Management
Yale School of Forestry, New Haven, Conn.

THE first edition of this valuable treatise appeared in January 1942. The newly published revision attests well to the thoroughness and exactness of its predecessor, since scarcely a correction was found necessary in text, formulas, tables, or figures. One new chapter has been added, titled "Double sampling of individuals in representative sampling of groups; systematic computations." The authors must again be commended for assembling in one volume the variety of sampling techniques most appropriate to the complex fields of forestry and range management and for suggesting many methods of approach to sampling procedures in cases where earlier and current "attempts to extract sampling error are more akin to the art of the conjurer than to scientific assay." The authors state their purpose to be that of encouraging the practicing forester or range technician to "acquire the art of planning—and executing—suitable sample procedures, such that (1) the real error may be assessed unambiguously; and (2) the best estimate is obtainable . . . consistent with the time and funds available for the sampling work." Mathematical derivation and proof is held to a minimum and involved technical terminology is restricted with the result that the text can be readily understood by the forester or range manager with only a preliminary knowledge of the statistical method. If any criticism is to be made of this book, it must be toward the restriction indicated by the title, for surely the methods advocated will find a far broader application than in the fields of forestry and range management alone.

The first part, consisting of two chapters, deals with the statistical background of sampling in its simpler aspects. The second part (Chaps. 3–7) deals with direct estimates by sampling and the third part (Chaps. 8–13) with in-

direct estimates through regression. A short appendix supplies some of the essential mathematical background.

Part II on direct estimates by sampling takes up in order typical sampling schemes, starting from the unrestricted random sampling in finite and infinite populations, then proceeds to stratified sampling, the simultaneous sampling of two or more populations, subsampling and finally the representative sampling of irregular blocks of known or unknown area. Each of these has its appropriate place in the field of forestry. The forester may regret the scant attention that is given to systematic sampling, to which he has been addicted ever since the beginning of the use of sampling procedures in the estimation of wood volumes and which he probably will never discard completely. He knows that systematic sampling, especially with the double stratification that he uses (one for unit areas, the other for forest types) gives good results and he is supported in this view by the few known investigations involving the checking of 100 per cent enumerations by various types of sampling schemes. These have shown that systematic sampling gives results of high precision, superior to random or stratified random sampling under at least several typical conditions, but a precision which cannot be proven on the basis of the data and under current lack of a suitable mechanism of analyzing systematic samples. Since the publication of the first edition of this book, several authors have investigated the case of systematic sampling, including Osborne, Yates, Madow and Madow, Finney and others, and have demonstrated its suitability under certain conditions. The advantages of systematic sampling in forestry are so many from the administrative and financial point of view (and this is probably true also of many other fields) that statisticians could well devote more time and effort in developing a suitable theory.

Part III deals with the particularly valuable tool of regression as an aid in indirect estimation. Starting with simple linear regression for cases where the independent variable is free from sampling error and where it is not, it leads to the utility of purposive and mechanical selection of samples in obtaining an efficient regression equation. Conditioned regressions and the use of weights in such regressions is followed by a short treatment of non-linear regressions. Regression in representative sampling is shown to be an effective device of correlating ocular estimates with measured values taken on part of the general sample area. There follows the new chapter of double sampling in the representative sampling of groups, a topic which is only briefly treated, but appears to be the introduction of a rather detailed discussion of the systematic computation of normal equations. The latter appears to be the real purpose of this new chapter. The final chapter points out certain practical aspects of sampling, including the definition of objectives; bias; size, shape and structure of sampling units; the character of the sample itself; sampling intensity; and allocation of costs. A final new section, which does not appear in the earlier version, handles the allocation of optimum sample size to different strata. This section is called for in view of recent develop-

ments in aerial photogrammetry, whereby an advance stratification of a forested area can be made on a map showing groups of timber of varying average volumes and variances.

Cybernetics. *Norbert Wiener* (Professor of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass.). The Technology Press. New York 16: John Wiley & Sons, Inc. (440 Fourth Ave.) and Paris: Hermann et Cie, 1948. Pp. 194. $3.50.

REVIEW BY SEBASTIAN B. LITTAUER
Associate Professor of Industrial Engineering
Columbia University, New York 27, N. Y.

CYBERNETICS is the science of communication and control in machines and men. The title was coined from the Greek χυβερνήτη in order to give identity to a new set of concepts developed by the author and his associates and further integrated by the author into a unified discipline as presented in this work. It encompasses a variety of branches of science in a manner which convincingly demonstrates the vitality and fruitfulness of the point of view of unification in the sciences. Many problems in neurology, psychopathology and communication engineering are shown to have a common core of meaning in which the methods of mathematics, logic and statistics are intrinsic and unifying factors.

The scope of the book is such as to appeal to a wide audience, including not only the specialists familiar with one or more of the particular fields dealt with, but also the non-specialist who is seriously concerned with the social implications of achievements in cybernetics. For, although the presentation is in considerable part mathematical, there is sufficient general discussion to acquaint and alarm any reader with the aims and potentialities of cybernetics. Of this the author takes cognizance, and in the closing pages of the introduction—an intensely interesting document on the circumstances which motivated and influenced these researches—he warns of the possible consequences to man that may follow from the creation of sensitive automata which can replace not only the human arm but also, in its simple functions, the human brain. In this, in spite of the fact that in the chapter on Newtonian and Bergsonian Time he argues the essential similarity between the functioning of the living organism and the cybernetic mechanism, he enjoins upon us the moral responsibility not to identify the human being as a commodity whose value is determined by the market place.

There are two fundamental aspects of the present work which largely contributed to its development as a new and unified science. One centers around the concept of the message and transmission of information and the other stems from the identification of the problems incident to the development of this concept in the nervous system with those encountered in some mechanisms. The common phenomenon of undesired hunting as a re-

sult of excessive feedback is quite analogous to a pathological condition known as purpose tremor. The recognition of this relation led to the study of certain aspects of neurophysiology as a feed back cycle of information circulating from the nervous system to the muscles and reentering the nervous system through the sense organs. The possibilities inherent in research from this point of view were presented by Wiener, Rosenbluth, and Bigelow in a paper entitled *Behaviour, Purpose and Teleology* which appeared in Philosophy of Science in 1943. The similarity of the information cycle in man and mechanism is suggested as a fruitful means of study of the former by experiment with the latter.The promise of this work presents a strong argument for combined effort by workers in each of these fields, fortified by some common ground of knowledge and vocabulary. It is part of the mission of cybernetics to encourage this closer tie.

The message—to quote—"is a discrete or continuous sequence of measurable events distributed in time—precisely what is called a time series by the statisticians." An operator or apparatus—predicitor—which follows a message finds conflict between the necessity for fidelity of response to smooth inputs (slow rates of change) and rapid response to sudden and large changes of input. Reconciling this conflict requires an appeal to the statistics of time series and the calculus of variations in order to find an operator deemed one of optimum prediction in that the mean square error of prediction is minimized.

The problem of disentangling a message from contaminating noise, or for that matter separating two messages, encountered in wave filter design presents a similar statistical picture. The transmission of information is statistical in that a prediction operator, in some specified sense optimum, must be based on the statistics of the time series of the message to be followed. The transmission of information is the transmission of alternatives, where the unit of information is that transmitted as a single decision between two equally probable alternatives. The amount of information is expressed as the negative logarithm to the base two of the number of such decisions that are to be made in attaining a particular observation. This it can be seen is the negative of the measure of entropy whence information or message is identified with negative entropy or the degree of organization in a system. The system of statistics developed for handling these problems becomes that of knowledge which can be expressed in a binary system. The results of these methods in determining prediction operators have been applied to computing machines, wave filters, simulated nervous system information cycles and the like, and found to be practically effective.

The mathematical developments are given in the chapters: Groups and Statistical Mechanics; Time Series, Information and Communication; Feed-Back and Oscillation. Familiarity with Lebesgue integration, probability theory and Fourier analysis, as well as with electrical circuit theory is helpful in reading these chapters. Nevertheless, their essence is remarkably well presented in verbal form. The statistics presented are based on knowledge of

the complete past of a time series; prediction based on sampling the history of a time series has but recently begun to be developed by a number of workers.

Implications and consequences of the theory developed are numerous. One possibility in quantum mechanics is quite promising; for, granting a hypothesis on the state of cosmic noise and degrees of freedom of the system the author's theory of information implementing the concept of negative entropy may be the proper means for deriving the Schroedinger equations from the Maxwell equations. It will be interesting to observe developments in this direction.

A striking feature of this book is the highly provocative nature of some of the author's speculations. For example, anyone familiar with problems in psychiatry and related attempts at practical therapy will be strongly persuaded by the cybernetic explanation offered for the nature of a class of disturbance and for the occasional effectiveness of its current treatment. Certainly within the field itself no explanation has met acceptance, and the stimulating suggestions offered by the author are to be welcomed for their promise of effective and practical progress. To the largest body of readers of cybernetics, and to the readers of this journal in particular, the chapter, Information, Language and Society, speaks most directly. The author is not optimistic about the immediate usefulness of the methods of cybernetics in the resolution of the problems of society, as are some of his colleagues in the fields of anthropology and sociology who have been urging him to take off in their directions. But he does show quite simply and directly in terms of the ideas developed in this book that most of the apologetics for the state of society in which our literature abounds, can in the very nature of things be but abortive and sterile efforts. In a framework within which there is no statistical control and in a time scale in which there can be only short runs, there can be no science of society, if science means prediction. The direction for fruitful work in the study of society is definitely pointed. If the influence of this book, and in particular of the last chapter, is such as to bring about a more realistic consideration of the problems of human survival, it will have done well.

Cybernetics had to be written—not only for the formal science which it presents—but also for the stimulation and enlightenment it offers to the non-scientist who in his turn does influence the direction of scientific inquiry. This is a unique book in its content—but it is also exceptional for its style, which manifests at times an expressive grace and incisiveness which cannot but compel the enthusiastic attention of the reader. Since it is reasonably certain that there will be further editions, a few improvements will be welcomed, such as correction of a number of misprints in the mathematical and verbal text, and inclusion of an index, as well as more references to the literature. In its depth, breadth and unity, Cybernetics is a powerful and important work; the author is to be congratulated for bringing his scientific knowledge and insight to bear upon problems of profound concern to a wide audience.

Sampling Methods for Censuses and Surveys. *Frank Yates* (Head of the Department of Statistics, Rothamsted Experimental Station, Harpenden, Herts, England). London: Charles Griffin & Co., Ltd., 1949. Pp. xiv, 318. 24s. (New York: Hafner Publishing Co., Ltd. $6.00.)

REVIEW BY W. EDWARDS DEMING
Adviser in Sampling, Bureau of Budget
Washington

HERE at last is a real book on sampling. It is a pleasure to review such an outstanding contribution. In the hands of an expert a probability sample can be designed so that it meets pretty closely some prescribed tolerance of sampling error (such as $\pm 2\%$, $\pm 15\%$), with a desired probability, and at the lowest possible cost per unit amount of information. Moreover, regardless of any assumptions that went into the planning, any estimate made from the sample is accompanied by an irrefutable index of precision. All this is made clear. In modern sampling practice it is also considered desirable to procure ancillary information concerning various alternative procedures, such as a different version of the questionnaire, a different plan of hiring, training, or interviewing, or supervising. This ancillary information illuminates the results obtained from the main survey, and is obtained by carrying out simultaneous or supplementary samples in addition to the main survey. Such ancillary information is particularly desirable for large surveys and complete censuses, as is coming to be the practice.

A sufficient background of mathematical theory and practical experience in dealing with human populations, farms, agricultural production, inventories, business activity, and numerous physical materials, have now accumulated by which probability samples may now be designed in a great variety of materials and problems.

The techniques of sampling have advanced rapidly during the past few years, but the requisite knowledge and experience have been largely confined to an inner circle of masters who have trained numerous apprentices by the spoken word. This is an unhealthy state of affairs, because the demand for competent statisticians has been running further and further ahead of the supply. No book can ever replace the privilege of working with a master, but Yates, indeed one of the masters, has put about as much inspiration and guidance into a book as is possible. Unfortunately, there is no royal and easy road to the top, but statistical teaching centres will now have an authoritative book to use for studies in sampling. The book will be useful for instruction in internships, and the private study of innumerable struggling, mathematically-inclined statisticians in government offices throughout the world will now be much more effective.

Other study and teaching aids in advanced theories of sampling do eixst; for example, Thionet's *Méthodes Statistiques Modernes des Administrations*

Fédérales aux États-Unis and *A Chapter in Population Sampling* produced by the sampling staff of the Bureau of the Census in Washington, 1947; also Mahalanobis's unsurpassed article entitled "On Large Scale Sample Surveys," in the *Phil. Trans. Royal Soc.*, Vol. 231B, 1944; and Walter Hendricks's *Mathematics of Sampling* (Virginia Agricultural Experiment Station, 1948), but these productions have not the scope of Yates's book.

The very title of the book is a clever statement of the attainment of modern methods of sampling. The title suggests, as is true, that sampling methods are now used not only for occasional or periodic special surveys of various kinds, but actually for replacing, broadening, and calibrating the data heretofore reserved for complete censuses of population, agriculture, commerce, etc.

The fact is that sampling is the modern approach toward all kinds of data. Sampling is the art and science of acquiring whatever information is desired, at the lowest possible cost, and with an objective index of precision. For detailed and precise tables by small areas, the most economical sample may of course turn out to be a complete census.

A glance at the table of contents is sufficient to show the extremely broad coverage that the author has included. Space will not permit a list of the topics treated, but they include a discussion of the ordinary biases that are encountered and the ways in which modern sampling design avoids and corrects these biases. The book begins at the beginning—the definition of the sample unit or units that are to be used, and the building or acquisition of the frame. In accordance with the recommendations of the U. N. Sub-Commission on Statistical Sampling, the term *frame* denotes a clear and unambiguous listing or mapping of the sampling units. In multi-stage sampling a frame will be required for every sampling unit that falls into the sample, in preparation for the next stage of sampling. The author discusses the efficiencies of various ways of drawing various kinds and sizes of sampling units, and of calculating the estimates.

In modern sampling design, the formula and procedure by which the estimates are to be prepared are as important as the procedures of selection; in fact, the two together constitute the sample design. Yates gives a splendid treatment of various methods of computation, along with excellent discussions of the amount of labor that is saved or entailed through the use of constant and variable weighting factors. He gives a well-balanced treatment of ratio-estimates, two-stage sampling, stratified sampling, variable sampling fractions, for some of which he was required to develop new theory to fill in the gaps. The estimate of the gains arising from multiple stratification (Sec. 8.4), and of using a plan of partial replacement of sampling units in successive samples (Sec. 88) will illustrate the breadth of techniques that are covered. A simple cost-function is treated on pages 283ff. In teaching the book, it might be well to make clear to the students that there is a cost-function associated with every sample design, and that sample-design in the

modern sense means using whatever plan will produce the most information per dollar.

The illustrations are mostly drawn from the author's wide experience in surveys of agricultural and wood lands, but this is excellent, because the fundamental principles are the same regardless of the material sampled.

The reviewer is in full agreement with the statement on page 59 that where a full census is compulsory, a sample should also be compulsory.

The well-chosen bibliography at the back, divided into convenient sections, will be very useful.

This review can only conclude with thanks and congratulations to the author: his book heralds a turning-point in statistical history.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

CONTENTS OF VOLUME 44

INDEX TO VOLUME 44: 1949

ARTICLES

BOOK REVIEWS